URBAN DYNAMICS: LONGITUDINAL CAUSAL RELATIONSHIPS AND FUTURE TIME SERIES FORECASTING

by

Faizeh Hatami

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Geography

Charlotte

2023

Approved by:

____
Dr. Jean-Claude Thill

____
Dr. Michael Ewers

____
Dr. Wei-Ning Xiang

____
Dr. Rajib Paul

ABSTRACT

FAIZEH HATAMI. Urban Dynamics: Longitudinal Causal Relationships And Future Time Series Forecasting. (Under the direction of Dr. JEAN-CLAUDE THILL)

Studying urban dynamics is essential given the ever-increasing changes in urban areas with all its ensuing consequences, whether negative or positive. It is of paramount importance to take into account the temporal dimension of urban dynamics when studying its patterns and processes. Nevertheless, the majority of studies overlook this consideration and take cross-sectional research approaches. Moreover, a large body of literature in urban dynamics is dedicated to the explanatory analysis and causal inference only, neglecting the importance of predictive analysis. Addressing these two main gaps, this research explores urban dynamics through both causal inference and predictive modeling using longitudinal research designs. Urban dynamics are studied from two aspects in this work; transportation/land-use interactions, and economic growth. In the first article, the impact of built environment on commuting duration is assessed in 2000 and 2015 in Mecklenburg County, NC using spatial panel data models. Results show that the built environment has a statistically significant impact on commuting duration. However, it is important to note that the practical magnitude of the impact is small. In the second and third articles, the business performance of businesses are forecasted for non-business services and business services respectively in Mecklenburg County, NC, using recurrent neural networks long short-term memory deep learning method. After building and training the sequential model, its predictive performance is assessed using out-of-sample evaluation.

ACKNOWLEDGEMENTS

First of all, I would like to thank and express my gratitude to my committee chair Dr. Jean-Claude Thill without whom completion of this research and my Ph.D. education was impossible. Second, I would like to thank my other committee members; Dr. Michael Ewers, Dr. Wei-Ning Xiang, Dr. Harrison Campbell and Dr. Rajib Paul. I sincerely appreciate all my committee member's support and input which helped the development of this dissertation.

In addition, thanks to all my colleagues including Dr. Paul Jung, Dr. Behnam Nikparvar, Dr. Mohammad Mokhlesur Rahman, Dr. Daniel Yonto, and Dr. Daidai Shen who made this journey a wonderful and fruitful experience for me.

Last but not least, special thanks to my husband Ahad Pezeshk Poor for all his support during this long journey. Many thanks to my parents and sister to whom words are not enough to express my gratitude.

TABLE OF CONTENTS

Contents

LIST OF TABLES

TABLE OF FIGURES

LIST OF ABBREVIATIONS

DL	deep learning
ML	machine learning
TOD	transit-oriented development
CTAD	compact and transit-accessible development
CBD	central business district
KIBS	knowledge-intensive business services

CHAPTER 1: INTRODUCTION

As complex systems, urban areas are of high importance to study. They have been increasingly becoming the population centers of the world as in 2020, in the United States, 82.7% of the population lived in urban areas. This urban population is projected to grow to 89.2% by 2050 (World Bank, 2021). Being focal points of human capital, urban areas are big concentrations of economic activities due to economies of scale. As a result of economies of scale or agglomeration economies, firms and economic entities benefit from concentrating near each other by reducing their production or distribution costs, and by knowledge spillovers. Urban areas provide great opportunities for economic entities to agglomerate and increase their productivity (Rosenthal & Strange, 2004). On the other hand, large concentrations of humans, economy, flows of goods and services and interactions between humans and the environment in cities have led to social, economic, and environmental problems, which makes them more important to study as complex systems.

Urban areas have been considered complex systems since the 1960s when Bertalanffy (1969) introduced applications of systems theory to other disciplines, such as social sciences. Complex systems comprise elements or subsystems and interactions between them. In an urban context, these subsystems include land uses indicating different types of economic activity and residential locations within the context of the environment making the nodes of a structure in the form of a network. These elements are connected through network edges such as flows of goods and services, input-output linkages between different economic sectors, linkages between supply and demand, and linkages between production and consumption. In complex systems, the relationships between subsystems are nonlinear. This means that the system responds differently to a variable at different states. These systems are complex as understanding the behavior of the system or its elements is not simple. They are adaptive, meaning that systems can learn from patterns and processes and react over time (Batty, 2009; Ladyman et al, 2013). Given the mentioned properties of urban systems, different theoretical approaches and models have been developed to understand the structure and relationships between the subsystems.

1.1.Studying cities as dynamic complex systems

Many of the theoretical approaches that were developed to explain urban systems are static, while cities are dynamic phenomena. These approaches aim to investigate relationships between urban elements by simulating them in a cross-section in time (Batty, 2009). Many of these approaches, mainly in the group of location theory, including the aggregate theory of social physics based on energy and potential, the aggregate theory of macroeconomic interactions between production and consumption, and the disaggregate theory of microeconomics of competing land uses, are based on an equilibrium status assumed in cities (White et al., 2015; Wilson, 1998). Using fixed equations, they assume that the relationships between variables are stationary over time (Allen, 2000). One issue with this static point of view is that heterogeneity in urban phenomena is an important concept; however, when a system is studied in shorter times, such as a cross-section in time, the heterogeneity will be removed or reduced as a result of averaging. This issue has been taken into account in urban studies, and urban systems theories and models have evolved towards including the dynamics of changes in simulations (Batty, 2009). Examples of these approaches are agent-based modeling and cellular automata, which were introduced to urban studies in the 1980s (Chen, 2012; Pinto & Antunes, 2007). These models take a bottom-up approach to studying complex patterns using interactions between individuals internally (Irwin & Geoghegan, 2001). Similarly, the heterogeneity issue as a result of averaging arises over space too. Many urban models study the urban dynamics in large-scale aggregates, assuming that relationships are stationary over space. Considering the mentioned shortcomings, studying urban phenomena in smaller spatial and temporal resolutions helps the model to capture the heterogeneity and dynamics.

1.2. Components of urban dynamics

Given the importance of the dynamic dimension of urban phenomena, a large body of literature has been developed in urban studies and urban planning to study urban dynamics (Batty & Sun, 1999; Capello & Nijkamp, 2004; Forrester, 1970; Hunt et al, 2005; Ramachandra et al, 2012; Pumain & Reuillon, 2017; Wegener, 1994). As urban systems are composed of various subsystems, urban dynamics studies have focused on the relationships between two or more subsystems. Some models, namely unified models, aim to capture the overall structure of the system incorporating the subsystems in an integrated way. Some others, namely composite models, consider the system as a hierarchy of related but autonomous subsystems (Wegener, 1994). With the majority of modeling frameworks being in the group of composite models, urban dynamics studies have studied the phenomena in different areas. A large number of studies have been on land use and land cover change (Batty, 2005; Cabral & Zamyatin, 2006; Herold, 2005; Li et al, 2015; Puertas, 2014; Sharma et al, 2012; White et al, 2000). These models usually use remote sensing raster data to study the changes in land use and land cover and predict urban growth in the future. They often utilize models such as cellular automata and agent-based modeling to simulate land use and its changes.

Another important area of study in urban dynamics is modeling the interactions between land use and transportation (Acheampong & Silva, 2015; Eboli et al, 2012; Kii et al, 2019; Pfaffenbichler et al, 2010; Wegener, 2021; Wegener & Fürst, 2004; Wilson, 1998). Different land use types, including residential and non-residential types such as industrial, commercial, retail, agricultural, and recreational, mainly indicate that economic activities are considered as interacting nodes and transportation as a proxy for flows of goods, humans, services, and knowledge between them. The relationships between transportation and land use have been studied in both ways and the importance of integration of the two in planning policies and interventions is well understood. Some studies investigate the impact of land use and its properties such as residential and employment density, composition or land-use diversity and land value on travel behavior of residents, workers, consumers and visitors (Cao & Fan, 2012; Ewing & Cervero, 2010; Handy et al, 2005; Hong et al, 2014; Leck, 2006; Ma & Cao, 2019; Wang & Zhou, 2017; Zhang, 2012). Others have studied the impact of transportation policies on land use as the location behavior of residents and firms (Leitham et al, 2000; Nurlaela, & Curtis, 2012; Zhu, 2013; Zondag & Pieters, 2005). In the land use-transportation integration studies, a large body has been dedicated to studying rail transit including subway metro and light rail as one of the most important transportation policy investments in the United States. Its impact has been investigated on economic development (Crampton, 2003; Credit, 2018; Knowles & Ferbrache, 2016; Park, 2018), travel behavior (Boarnet et al, 2013; Kim, 2007;

Park et al, 2018) and property values (Debrezion et al, 2007; Diaz & Mclean, 1999; Hess & Almeida, 2007; Yan et al, 2012).

Another important dimension of urban dynamics that has received a great deal of attention is the dynamics in economic development. A large number of these studies have investigated the changes in spatial locations of employment centers influencing the overall structure of the city. Urban areas mainly have been experiencing an evolution from a monocentric structure to a polycentric structure (Boarnet et al, 2017; Brueckner et al, 2001; Gaschet, 2002; McMillen, 2001, McMillen, 2003, Sarzynski et al, 2005). In the traditional monocentric cities, employment opportunities were located in the Central Business District (CBD) influencing the other urban subsystems. Given the concentration of work opportunities in the CBD, land values increase with decreasing distance to the CBD. Different types of land use such as commercial, residential and industrial are structured around the CBD based on the land value and the rent they tend to pay, according to the bid rent theory (Alonso, 10960). However, in a polycentric city there are two or more employment subcenters. These subcenters have developed as they have less land cost for firms and less transportation cost for suburban residents (McMillen, 2001). They can be created in highway interchanges, in the periphery of transit stations or along the rail transit (McMillen, 2003).

One of the forces of creating these new subcenters is agglomeration economies or economies of scale, which is another important strand of literature of urban dynamics and economic development. According to agglomeration economies, firms and businesses tend to locate themselves close to each other and cluster as they gain some advantages of this clustering. Some forces leading to agglomeration economies are input-output linkages between industries resulting in lower costs, closer proximity between firms resulting in lower transportation costs, labor pooling and access to the skilled workers, knowledge transfer, consumption externalities such as comparison shopping and multipurpose shopping behavior, and lastly, natural advantages such as access to consumers and access to transportation infrastructure (Billings & Johnson, 2016).

However, one issue in the urban dynamics literature is that most of the studies are in coarse-granularity regions or systems of systems. Scale plays an important role in results of studying spatial phenomena due to the Modifiable Areal Unit Problem (Openshaw, 1979). In order to address this gap, this study investigates urban dynamics at the higher resolution of census block groups representing neighborhoods. In addition, this study takes a longitudinal approach to study the phenomena over time instead of a cross-section of time.

1.3. Research questions

This dissertation studies urban dynamics from two interconnected dimensions; first, relationships between the built environment and transportation and second, economic development. In the first question, the impact of the built environment on commuting duration is evaluated and in the second and third questions, economic growth is predicted for particular industries using urban dynamics components such as built environment, sociodemographic and economic predictors. Industries are divided into two groups of non-business services and business services for questions two and three based on the 2-digit NAICS codes. These two groups are treated separately since the dynamics and the agglomeration forces vary across the two groups (Giuliano et al., 2019; Hanlon & Miscio, 2014).

Business services are different from non-business services because of their demand side, which usually encompasses firms or organizations. They are knowledge intensive. knowledge is both a production factor and a product of these services (Strambach, 2010). On the other hand, non-business services provide services to individuals and as a result, they are dependent on foot traffic and in person interactions with their customers. Economic trends in non-business services are more stable over time since they provide services to individuals and not businesses. As a result, they are less sensitive to economic status in other sectors while business services are strongly influenced by the magnitude of demand and dynamic processes of production in other industries (Manalo & Orsmond, 2013). Business services comprise a larger share of the economy in terms of count of employees and sales volume. Non-business services are more dependent on the neighborhood dynamics while business services are more dependent on input-output linkages and dynamics in other industries rather than the location. In addition, ranges of sales volume are much higher in business services than in non-business services. Their labor pool is very different with business services having high skilled workers and non-business services having low skilled workers.

For question one, this study evaluates the impact of the built environment on commuting duration. This part of the study is aimed to answer the question whether the built environment has an impact on travel behavior of residents. It is hypothesized that built environments with higher degrees of built environment variables, namely 5 D variables including density, diversity, design (Cervero & Kockelman, 1997), destination accessibility (Handy, 1993) and distance to transit (Ewing & Cervero, 2001) lead to lower travel consumption. This relationship is investigated in two steps. First, census block groups are classified into built environment types of exurban, suburban, urban and compact and transit-accessible development (CTAD) using Ward's clustering method (Ward, 1963). Variables related to the built environment, namely the 5 Ds, are used for clustering. Second, the impact of the built environment types on commuting duration is assessed in Mecklenburg County, NC, in two years of 2000 and 2015 as before and after years of establishment of light rail transit and development of TOD areas. The study is conducted at the census block group geographic level. This study is aimed at addressing two gaps in this research area. One gap is the selection bias which means that the association between changes in built environment and changes in travel behavior does not necessarily indicate causality. In other words, the built environment may not necessarily lead to travel reduction, but individuals may self-select themselves into those built environment types as a result of their travel preferences. This study addresses this issue by employing a longitudinal design to remove unobserved factors such as people's travel preferences and attitudes. The second gap is the spatial dependence in urban socioeconomic phenomena which is not taken into account in this literature sufficiently. Using nonspatial and spatial panel data models, this study addresses self-selection and spatial autocorrelation issues.

For questions two and three, economic growth for particular industries in business and non-business service groups are predicted and evaluated for three years ahead in Mecklenburg County, NC, using historical data of the years 2002 to 2018. The business performance of neighborhoods is predicted using three indicators of employment, sales volume, and labor productivity for non-business services, and labor productivity for business services.

Deep learning is used for the forecasts. One of the objectives of this research is to address some of the gaps in urban economies literature. As the first gap, the urban economics literature mostly investigates the causal relationships between its components. However, prediction of economic growth is of high importance as it can be used for understanding future changes in spatial distribution of employment centers, and for identifying growing and declining neighborhoods in advance. As another shortcoming in the literature, the majority of urban economies studies are in large scales of cities, regions, or nations (Billings & Johnson, 2016). Nonetheless, by studying the economy at the higher resolution of the neighborhood, particularly, forecasting will provide us with prospective insights beneficial to a wide range of stakeholders such as planners, business owners, developers, residents, and investors.

Mecklenburg county is selected as the study area of the three chapters because it has been very dynamic over the last decades in terms of its economy, demography of its population, and transportation-land use interactions. Placing the city of Charlotte as its seat, Mecklenburg county has experienced a fast population (U.S. Census Bureau; 2021a; U.S. Census Bureau; 2021b) and economic growth (LEHD, 2021) over the recent decades. Anticipating the mentioned growth in the early 1990s, transportation-land use plans have been developed to enhance the connectivity of functional areas of the county such as employment centers through public transit infrastructure in five corridors (City of Charlotte, 1998). In these plans, compact and mixed uses have been proposed for areas along the transit corridors to place a variety of economic activities.

In the next chapters, each of the three research questions is discussed in further detail, their existing literature is reviewed, and the current research gaps are discussed. Then, research frameworks are proposed and data from Mecklenburg county as the study area is used to investigate the research questions.

CHAPTER 2: SPATIOTEMPORAL EVALUATION OF THE BUILT ENVIRONMENT'S IMPACT ON COMMUTING DURATION

Abstract

Upward trends in commuting duration and distance due to urban sprawl in the United States has raised concerns about the ensuing environmental, social and economic problems. Various urban planning approaches have been developed hypothesizing that built environment variables such as density, diversity, design, distance to transit and destination accessibility contribute to reducing travel consumption. This study evaluates the impact of the built environment on commuting duration in Mecklenburg County, North Carolina, in two steps. First, the built environment is classified into four types of exurban, suburban, urban, and compact and transit-accessible development (CTAD). Second, the impact of built environment types on commuting duration is evaluated for 2000 and 2015 using spatial panel data models controlling for selection bias. Results show that CTAD areas have shorter commuting durations than other areas in 2015; however, the commuting duration in both CTAD and urban areas increased over time. Given the multifaceted nature of urban transportation-built environment interactions and their importance for sustainable futures, this calls for further attention of urban researchers and planners to more comprehensively consider the various dimensions of this matter with an explicit focus on the changing nature of urban environments.

Keywords

Commuting duration, built environment type, D variables, spatial panel data models, selection bias, urban mobility

2.1. Introduction

In the United States, there has been an upward trend in urban sprawl development and car dependency (Barrington-Leigh & Millard-Ball, 2015), in average vehicle miles traveled (VMT) (US Department of Energy, 2019) and in commuting duration (American Community Survey, 2019) over the recent decade. These trends have awakened concerns about the ensuing environmental, economic and social costs, such as fossil fuel consumption, environmental pollution, climate change and social exclusion. In order to alleviate these problems, various urban planning and design approaches have been developed. These include new urbanism, smart growth, sustainable urbanism, compact development and Transit-Oriented Development (TOD). Their aims are to reduce travel demand and consumption by bringing destinations closer together, to improve accessibility and increase the number of travel options for a broad variety of social groups. The typical argument is that these objectives can be achieved by increasing residential and employment density, enhancing diversity of social groups, housing types and affordability, land-use mix, improving urban design with the development of pedestrian and bicycle-friendly neighborhoods with small block sizes, and increasing accessibility to a variety of transit options.

Whether or not these approaches have been effective at improving travel outcomes and related issues discussed above has been germane to the ongoing urban policy and planning debates situated at the intersection of urban sustainability, urban livability and quality of life, and community-based approaches to city building, as advocated by Jane Jacobs (1961). If travel demand impacts are proved to exist, policy makers can more readily feature the relevant urban planning and urban design principles as part of an integrated vision and practice for transportation planning. As a result, a large number of studies have investigated the relationship between the built environment and travel behavior over the recent decades (e.g., Crane, 2000; Ewing & Cervero, 2001; Ewing & Cervero, 2010; Stevens, 2017). These studies have been aimed at finding an answer to whether changes in the built environment have impact on travel demand and consumption. Testing this hypothesis in different study areas in the United States using different data sources and different methodologies, studies have found somewhat inconsistent results. Some studies concluded to the existence of positive impacts (Cao, Mokhtarian & Handy, 2007; Cervero & Murakami, 2010; Chen, Gong & Paaswell, 2008; Izanloo et al., 2017; Khattak & Rodriguez, 2005; Salon, 2015; Zhou & Kockelman, 2008), but some others found either no statistically significant relationship or very weak impacts (Bagley & Mokhtarian, 2002; Crane & Crepeau, 1998; Etminani-Ghasrodashti & Ardeshiri, 2015; Ewing & Cervero, 2010; Nasri & Zhang, 2015; Stevens, 2017; Wang, 2013).

Many of the studies investigating the relationships between the built environment and travel behavior suffer from some shortcomings in research design and methodologies, which may have led to the mentioned confounding findings. One of these methodological issues is the selection bias or self-selection (Boarnet, Crane, 2001; Cao, Xu & Fan, 2010; Ewing, Cervero, 2010; Mokhtarian, Cao, 2008). Self-selection means that individuals' travel behavior may not necessarily be influenced by the built environment, but they may select themselves into those built environments due to their travel preferences. This is fundamentally an endogeneity issue that is known to be rather pervasive in social sciences (Ghose, 2019). The second issue that needs to be taken into account is the spatial dependence or autocorrelation that may explain a significant amount of variation in the outcome variable. Spatial autocorrelation means that there is a spatial relationship between values of geographic unit areas and their neighbors based on methods like geographic adjacency or proximity (Cliff & Ord, 1970). Positive spatial autocorrelation indicates similar values between close neighbors and negative spatial autocorrelation indicates dissimilar values between neighbors. Such spatial effect is often intertwined with internal and external contextualization considerations that may mask relationships (Thill, 2020).

Furthermore, it is important to distinguish the work trips and nonwork trips when studying the different components of travel behavior as their dynamics are not the same. In this study, the trip to work is selected as it is the trip purpose with the greatest structuring effect on other travel aspects for most individuals. Duration of work travel is taken as the dependent variable as it reflects choices made by individuals in terms of home place, work place, mode of travel and routing, as well as the reciprocal adjustments from the urban systems in the form of traffic congestion in dense urban areas. In addition to the commuting duration component of travel behavior, to measure the built environment, a variety of indicators including housing density, road density, intersection density, land-use mix, single family housing size, multifamily housing percentage, jobs-housing ratio, jobs accessibility and rail transit proximity, are used for measuring the multiple facets of the built environment and its context.

In this article, we revisit the relationship between the built environment and commuting duration through an empirical case study in the United States. To address the two methodological issues mentioned above, the study uses a longitudinal design as a means to control for self-selection in testing the impact of the built environment on commuting duration component of travel behavior. This impact is evaluated in two years

(2000 and 2015) in Mecklenburg County, North Carolina, block groups. By taking the differences between these two years in this longitudinal design, the selection bias is removed as an unobserved factor. In addition, spatial panel models are applied to account for the spatial dependence that may be embedded in the concepts, relationships, and data used for this analysis.

The study is conducted in two main steps to test the hypothesis that the built environment type has an impact on commuting duration and that built environment types with higher levels of density, diversity, design, destination accessibility, and transit proximity have shorter commuting duration. First, variables related to the built environment are used to classify the built environment in 2015 in four types comprised of the exurban, suburban, urban and compact transit-accessible development settings, with Ward's clustering method. Second, non-spatial and spatial panel data regression models are estimated to study the impact on commuting duration.

2.2.Literature review

There is a large body of literature on the relationship between the built environment and travel behavior that dates back to the 1990s. A majority of these studies found that the features of the built environment significantly impact on reducing travel demand and consumption while others found no evidence of such impact. When a relationship is detected, it is normally found that travel demand is lower where urban development assumes denser and more compact forms. Yet conclusions of research studies remain somewhat at odds. The disparity in conclusions reported in this literature may reflect less the true nature of this relationship than other design considerations such as the multidimensional and contextual nature of the relationship or its bidirectionality. Both the built environment and travel behavior are complex and multifaceted concepts. Considering their components, both concepts have been apprehended through a variety of dimensions across the extant literature.

The built environment has been apprehended by the so-called 5 Ds, which stand for density, diversity, design (Cervero & Kockelman, 1997), destination accessibility (Handy, 1993) and distance to transit (Ewing & Cervero, 2001). Density is measured by population, residential or employment density. Diversity is measured by land-use mix or diversity in housing affordability. Design is measured by intersection density or road density as a proxy for small block sizes, and therefore the ease of non-motorized travel. Destination accessibility is measured by the ease of access to jobs, retail outlets and services. Finally, distance to transit is measured by the distance to the public transit options. Some studies have directly examined the impact of these variables on travel behavior (Cervero & Kockelman, 1997; Crane & Crepeau,1998; Ewing et al., 2015; Handy, Cao & Mokhtarian, 2006; Nasri & Zhang, 2012), while some others investigated the heterogeneity in travel behavior in different built environment types defined based on the D factors mentioned above (Bagley & Mokhtarian, 2002; Handy, Cao & Mokhtarian, 2005; Khattak & Rodriguez, 2005; Salon, 2015; Zhou & Kockelman, 2008). As for travel behavior, it is standard to measure it through its components of trip length (distance or time), trip frequency and travel mode for different trip types. The working hypothesis is then that the prevalence of the 5 Ds that portray the built environment bring about a reduction in travel activities. Literature can be further synthesized as follows.

Using travel diary data in the San Francisco Bay Area, Cervero and Kockelman (1997) investigated the impacts of density, diversity and design on trip rates and mode choice and found a statistically significant response to these three built environment factors, although the effects were weak. Cervero (2002) used travel and land-use data from Montgomery County, Maryland, and found significant evidence that density and land-use mix had an impact on travel mode choice, while design had modest impact only. Cervero and Duncan (2006) studied the impact of jobs-housing balance and retail-housing mix on VMT with travel survey data from the Bay Area, California; they found a significant impact on travel reduction, with the jobs-housing balance being more influential. Using travel diary data from New York Metropolitan Area, Chen et al. (2008) studied how mode choice in work trips responds to population and employment density, accessibility and transit proximity in both home place and work place. They found that job accessibility had the most significant impact, followed by density and transit proximity. Zhang et al. (2012) investigated how factors of density, land-use mix, block size and CBD proximity influences VMT reduction using household activity survey data in the four metropolitan areas of Seattle, WA, Richmond-Petersburg and Norfolk-Virginia Beach, VA, Baltimore, MD, and Washington, DC; they found significant impact, but there was heterogeneity in this effectiveness both between and within metropolitan areas. Finally, using household travel data in 15 US regions, Ewing et al. (2015) investigated the impact of the 5 Ds on several travel behavior components, namely car trips, walk trips, bike trips, transit trips and VMT. They too concluded that significant impacts existed in the form of a drop in travel demand.

The other strand of literature boils the diverse characteristics of the built environment to a typology that can be more effective at discriminating the behavioral responses. Khattak and Rodriguez (2005) used travel survey data from Chapel Hill and

Carrboro, North Carolina, and found that built environment types including neo-traditional and conventional neighborhoods significantly deflate trip frequency by trip type. With travel survey data in California, Salon (2015) studied the heterogeneity in the relationships between the VMT and built environment characteristics in the built environment types (rural, rural-in-urban, suburb, urban and central city) and found that this relationship is strongly dependent on both the built environment type and trip purpose. In other words, for specific trip purposes and built environment types there is a significant impact, but not for others. In their study of travel survey data from 8 neighborhoods in Northern California, Handy et al. (2005) found significant differences in travel behavior between traditional and suburban neighborhoods. In a similar article using survey data from four traditional and four suburban neighborhoods in Northern California, Cao et al. (2007) found that changes in the built environment are significantly associated with changes in driving reduction, with a large negative impact from accessibility variables. Finally, Cao et al. (2010) studied the impact of residential location (namely, environments defined as urban, suburban, exurban and inner ring suburbs) on vehicle miles driven using regional travel diary data from Raleigh, North Carolina, as a case study, and found a significant impact with decreasing auto dependence and vehicle miles driven in areas that are more urban.

Studies mentioned earlier in this section found some relationship between the built environment and travel behavior in the sense that the 5 Ds are effective at reducing travel demand, either in all or in part of the built environment and travel behavior components. However, a small number of other studies found no such relationship or a rather weak relationship between the two. Among these, Crane and Crepeau (1998) studied this relationship on travel survey data in San Diego, California, and found evidence that design would have an impact on travel decisions but found little impact of density. Bagley and Mokhtarian (2002) studied the impact of residential neighborhood types (specifically traditional and suburban neighborhoods) on travel demand measurements of vehicle miles traveled, transit miles traveled and trip frequency. Using travel survey data from the Bay Area, California, they found very little impact.

Given the literature's rather inconsistent findings, Leck (2006) performed a meta-analysis on the impact of urban form on travel behavior to draw generalizable conclusions. In this study, urban form was measured by residential and employment densities, land-use mix and street patterns. Travel components included VMT, vehicle hours traveled (VHT), vehicle trips, non-work vehicle trips, probability of commuting by automobile, transit, or by walking. Studying 17 prior case studies on this topic, this meta-analysis concluded that residential and employment densities and land-use mix had statistically significant impact on travel behavior and dampened travel demand, even after controlling for socio demographic attributes. However, street layout did not show a significant impact. Ewing and Cervero (2010) conducted another meta-analysis on more than 50 studies of the relationship between the built environment and travel behavior, updated some of them and obtained the effect sizes. Their paper found that VMT was associated with accessibility and street network design, that walking was associated with diversity, design and accessibility, and that transit use was associated with transit proximity, design and diversity. It is important to point out that magnitudes of influences were found to be small. In a more recent Meta-analysis study of this topic, Stevens (2017) reviewed the literature to answer the question whether the built environment and compact urban development reduce travel, or conversely whether this link is inexistent or tenuous and other urban scenarios explain the evolution of the urban system. He discussed the selective reporting bias as an important issue and as one of the most importance potential reasons for the incongruous results in the literature. Studying the findings of 37 papers, he found that some of the D variables measuring the built environment had statistically significant impact on reducing VMT; however, the impacts were so small that he suggested planners to not rely on compact developments for reducing travel unless the costs were very small or their expectations of reducing travel were very low. In his meta-analysis work, the statistical significance and direction of impacts on travel varied over different built environment variables. Job accessibility by transit and jobs–housing balance had no statistical significance. Increase in population density reduced travel; however, increase in diversity and decrease in CBD proximity increased VMT. Similarly, Guerra (2014) found that areas closer to the CBD had longer VMT. Also, Jin et al. (2017) studied the impact of densification on travel duration and found that densification increased traffic congestion and consequently increased travel duration.

Aside from using somewhat different sets of attributes of the built environments and of travel behavior in their research design, these studies may also exhibit discordant conclusions owing to some methodological considerations, particularly how self-selection is addressed, if at all (Cao, Mokhtarian & Handy, 2007; Handy, Cao & Mokhtarian, 2005; Khattak & Rodriguez, 2005; Mokhtarian & Cao, 2008; Salon, 2015; Zhou & Kockelman, 2008). The self-selection issue boils down to the possibility that the association between the built environment and travel behavior found in the literature does not clearly indicate a causal relationship between the two. In other words, the built environment may not necessarily lead to changes in travel behavior (i.e. travel reduction), but it may well be the individuals' travel preferences that encourage them to select themselves into neighborhoods with specific characteristics (e.g. compact and diverse development with higher accessibility to transit and different modes of travel). Studies have used different approaches to tackling the self-selection issue, including direct surveying of individuals, statistical control through attitudinal information, instrumental variables estimation, purposeful sample selection designs, joint discrete choice models, structural equation modeling and longitudinal data models (Mokhtarian & Cao, 2008).

2.3. Methodology

This study is comprised of two parts for investigating the impact of the built environment on commuting duration. First, the city space is partitioned into four distinct types, namely the exurban, the suburban, the urban, and compact and transit-accessible development (CTAD), based on built environment elements. Built environment elements bundled together as built environment types are used instead of individual built environment variables to evaluate impacts on commuting duration to circumvent the potential collinearity among 5 D elements. Accordingly, we use a multivariate clustering approach to identify built environment types that denote the most discriminated structures in the built environment based on the combination of built environment variables. Practically, census block groups are classified in mutually exclusive and collectively exhaustive types through a stepwise process that incorporates Ward's clustering method (Ward, 1963). Second, we estimate a non-spatial panel data model of the effect of the built environment on commuting duration as well as several corresponding spatial models. In order to control for the selection bias issue, the difference in differences (DiD) technique is incorporated in our panel data models. The DiD method takes the difference in change in outcome variable in treatment groups from change in outcome variable in control groups as shown in equation 1:

$$y= \alpha + \beta_1 D + \beta_2 T + \beta_3 D*T + u. \quad (1)$$

In the above equation, *D* is a binary variable representing the treatment and control groups and *T* is a binary variable indicating times before and after the treatment; *y* is the outcome variable, *u* represents the error term and *α* is the intercept. The interaction term between the groups and time helps to resolve the selection bias issue by removing the unobserved variables. In this study, the two years of 2000 and 2015 are chosen as the before and after years of operation of the Lynx light rail service and of implementation of compact transit-accessible development along the transit corridor.

Commuting duration is selected to be studied as one of the components of travel behavior since work trips are the most important trip type for a majority of individuals. In addition, this variable effectively controls for traffic congestion that high density developments are expected to generate (Manville, 2017). This selection of dependent variable stands in contrast with many previous studies that used VMT or VMT per capita for investigating the relationship between the built environment and travel behavior; however, they are not as compelling as travel duration. As a matter of fact, as a travel distance measure, VMT does not account for traffic congestion in general, while the problem with VMT per capita is that the development of high density and diverse areas with high transit accessibility may decrease VMT per capita in the entire area but not necessarily in the compact areas.

2.3.1. Study area

The area of study is Mecklenburg County, North Carolina, and unit areas are census block groups. As the most populous county in North Carolina until 2020 and with the city of Charlotte as its seat, Mecklenburg County has been experiencing fast population growth and accelerated economic development over past decades. Anticipating continued growth of population and of the local economy, some plans were developed for this county to better integrate land use, urban design and transportation systems, such as the 2025 Transit/Land-Use plan which was finalized in 1992. This plan entails five transit corridors and developments along them; the first corridor started operating in 2007 as the Blue Line of the Lynx light rail service. Establishment of this light rail line has led to compact and diverse developments incorporating 5D design elements, including Transit-Oriented Development projects in proximity to the rail transit stations. In order to study the impact of the built environment types, two years of 2000 and 2015 are selected to capture the conditions before and after the establishment of the light rail line and the implementation of compact development areas.

2.3.2. Data sources

Data for classifying the built environment are collected from the National Historical Geographic Information System (NHGIS), including primary data drawn from the American Community Survey (ACS). Also, GIS shapefiles were sourced from the Mecklenburg County GIS center, while some travel and mobility data came from the Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES). Data from NHGIS facilitated the integration of data longitudinally, as this data source provides interpolated 2000 data in 2010 census boundaries. To control for the socio demographic attributes, data on median population age, median household income, median housing value, race, educational attainment and car ownership are collected from NHGIS and ACS. Commuting duration data are also collected from the ACS.

2.3.3. Built environment types: classification and factors

Partitioning and classification of the city space is done for the year 2015 in three sequential steps on the basis of built environment considerations. First, rural areas are extracted from the Mecklenburg County block groups using the Census Bureau's urban-rural classification data (US Census Bureau, 2010) to form the exurban area type for the purpose of the present study. Second, non-rural block groups are classified into either suburban or other environment using built environment factors of land-use mix, road density, intersection density, housing density, multifamily percentage and single-family housing size. Third, the latter set of block groups is classified into urban and CTAD using the built environment factors of land-use mix, jobs-housing ratio, jobs accessibility, housing density, multifamily percentage, with the condition that a block group must be within 0.5 mile of any Lynx Blue line station to be classified as CTAD. Table 1 reports the descriptive statistics of these variables in the pooled dataset. Ward’s method was used for steps two and three of the classification process.

Table 1. Descriptive statistics of built environment related variables in 2015 (N = 546 census block groups).

Built Environment Related Variable	Min	Mean	Standard Deviation	Max
Land-use mix (normalized entropy index)	0	0.001	0.001	0.009
Road density (ft/mi^2)	13,913	66,997	22,877.45	166,697
Intersection density (units/mi^2)	9.83	75.00	37.68	255.31
Housing density (units/mi^2)	32.7	1409.5	1211.7	12,959.3
Multifamily housing (%)	0	23.815	41.380	100
Average size of single-family lots (ft^2)	0	32,725	90,942	1,244,304
Jobs-housing ratio	0.005	2.032	9.583	191.076
Jobs accessibility (jobs)	0	772	1531	12,768

Further elaboration on the built environment variables used in this process is warranted. Housing density is the number of all housing units per square mile using ACS data sourced from NHGIS. Road density is the total length of street centerlines per square mile extracted from the street network shapefile from the Mecklenburg County GIS Center. For intersection density, four-way intersections are created from the street network shapefile and density is calculated as their count per square mile; this density is a proxy for block size, which tends to be smaller in more urban environment and CTAD neighborhoods. Land-use mix is calculated using the entropy index relative to a reference geography. The entropy index is a common method for measuring diversity in land use as a function of percentages of two or more land uses in an area. However, an issue in this method is that equal percentages of land uses will produce the highest degree of mixture. For example, 50% housing and 50% industrial land uses will create a perfect entropy index of 1 while it is not a desirable mixture. To solve, this issue, Song, Merlin and Rodriguez (2013) proposed to normalize this measure by using the percentages of land uses in a well-balanced reference geography, for example the entire study area, if considered well-balanced. To obtain the entropy index using the mentioned method, land use data are extracted from tax record dataset sourced from the Mecklenburg County GIS center. Land uses for the year 2015 were classified into the following categories: Commercial, Industrial, Recreation, Single-family, Multifamily, Institutional, Office and Other. After using this enhanced variant of the entropy index, one more improvement is made to the method. With the entropy index, larger block groups will have higher diversity scores since they can encompass a large variety of different uses. To circumvent this drawback, the entropy index is normalized by the area of the block group. Single-family housing size is the mean of the area of all single-family houses. Multifamily housing share is the percentage of multifamily housing of all residential units. The data for area of single-family lots and multi-family housing are obtained from the tax record dataset sourced from the Mecklenburg County GIS center. The jobs-housing ratio is the ratio of the number of employees to residents in each unit area using LEHD data. Jobs accessibility is the total number of jobs within the 5-mile distance of each areal unit, using LEHD data. Rail transit proximity is an indicator variable of whether an areal unit is within 0.5 mile of rail transit or not.

2.3.4. Non-spatial and spatial panel data models

A panel data model of commuting duration is estimated to test the impact of the built environment type, accounting for the selection bias. The model used in this study follows the Difference in Differences approach in that it takes the differences between both built environment groups and time periods into account. Two groups are included in the model specification: urban and CTAD; exurban and suburban serve as reference group. Second, interactions between these two groups and the year variable is added to the model to test changes in commuting duration in these groups over time. Socio-demographic variables are added to the model to control for their impact. Among the sociodemographic variables, the population percentage with some college degree, population percentage with Master's degree and median housing rent value are removed from the model due to the collinearity. Table 2 shows the descriptive statistics of the variables used in this model.

Table 2. Descriptive statistics of independent variables used in the panel data model of commuting duration (Minutes) (n = 1092 block groups for the pooled dataset, n= 546 for 2000 and 2015).

No vehicle Available (%)	**CBD Proximity (mi)**	**Ph.D. Degree Holders (%)**	**Bachelor's Degree Holders (%)**	**Associate's Degree Holders (%)**	**Median Population Age (Years)**	**Median Housing Value (Dollars)**	**Median Household Income (Dollars)**	**Housing Density (Units/mi^2)**	**Asian Population (%)**	**African American Population (%)**	**White Population (%)**	**Year**	**Independent Variables**
0	0.13	0	0	0	10.8	937.97	6791.47	1.66	0	0	0	Pooled dataset (2000 and 2015)	Min
7.58	7.39	0.78	22.74	5.77	35.42	189,209.30	60,983.93	1249.28	3.8	29.29	61.33		Mean
9.29	4.03	1.19	13.08	3.11	6.39	136,096.30	32,691.11	1102.41	5.14	27.55	29.3		Sd
64.8	19.41	13.14	57.98	21.08	80.5	1,169,800.00	250,001	12,959.33	54.87	99.33	100		Max
0	0.13	0	0	0	19.5	937.96	6791.46	1.65	0	0	0	2000	Min
7.16	7.39	0.81	26.56	6.61	34.23	161,953.90	56,414.17	1089.08	2.94	27.21	65.53		Mean
9.73	4.03	1.2	13.89	2.87	4.58	103,069.70	26,785.43	955.63	2.71	28.08	29.5		Sd
64.8	19.41	13.13	57.98	16.81	51.95	784,989.10	199,375.20	6402.19	18.92	99.32	100		Max
0	0.13	0	0	0	10.8	16,100.00	12,212.00	32.7	0	0	0	2015	Min
7.98	7.39	0.75	18.92	4.93	36.59	216,467.00	65,553.68	1409.47	4.66	31.37	57.12		Mean
8.8	4.03	1.16	10.95	3.11	7.61	158,008.00	37,149.00	1211.68	6.64	28.5	28.5		Sd
45.5	19.41	7.86	48.11	21.08	80.5	1,169,600.00	250,001.00	12,959.33	54.86	100	100		Max

After estimating the non-spatial panel data model, spatial autocorrelation in the dependent variable and in residuals is tested with Global Moran's *I* (Anselin, 1995) (see

Table 3). Based on Table 3, there is strong statistically significant autocorrelation in the dependent variable and statistically significant autocorrelation in the model residuals. As a result, both spatial lag and spatial error models are considered to handle the spatial dependences in the data (Anselin & Bera, 1998). Spatial lag accounts for the spatial dependence in the dependent variable and spatial error accounts for the spatial dependence in residuals in explaining variations in the dependent variable. In addition to the spatial lag and spatial error models, spatial Durbin and spatial error Durbin models are also implemented. The spatial Durbin model incorporates the spatial lag of the dependent variable and independent variables and spatial error Durbin model incorporates the spatial lag of the error and spatial lag of the independent variables in the model. All the spatial models are estimated using a spatial weight matrix based on the adjacency of block groups. This spatial weight matrix is an NxN matrix indicating whether two block groups are neighbors or not, with row standardized weights. N is the number of block groups. Queen contiguity criterion is used for determining spatial relationships between neighbors.

Table 3. Spatial autocorrelation in dependent variable and model residuals using Global Moran's I.

	Spatial Autocorrelation in Dependent Variable		**Spatial Autocorrelation in Model Residuals**	
Year	**Index**	***p* Value**	**Index**	***p* Value**
2000	0.751	0	0.33	0
2015	0.508	0	0.14	0

In addition to the Global Moran's I statistic, Lagrange Multiplier (LM) tests are used before running spatial models (Anselin, 1988) (see Table 4). LM-lag and LM-error test for spatial lag and spatial error in regression model, respectively. Robust LM-lag tests for spatial lag in presence of spatial error and robust LM-error tests for spatial error in presence of spatial lag. Based on LM tests, both spatial lag and error model are diagnosed.

All the spatial models are implemented in the Spatial Panels MATLAB package developed by Elhorst (2014).

Table 4. LM and robust LM tests for spatial lag and spatial error models..

Test	LM-Error	LM-Lag	RLM-Error	RLM-Lag
LM	79.3781	78.3621	4.9183	3.9023
p-value	0	0	0.027	0.048

2.4. Results

Figure 1 shows the results of the built environment type classification in 2015. Starting from the Bureau of the Census definition of urban and rural areas, 16 block groups were identified as exurban, mainly at the periphery of the county. Using the Ward's clustering method, 252 of the non-rural block groups are classified as suburban, 259 census block groups are identified as urban, mainly in the urban core and along the corridors of development. Finally, 28 census block groups are identified as CTAD. These CTAD areas are located in the CBD and along the south corridor, in proximity of the light rail Blue Line. Statistics of the centroid of each of these clusters are reported in Table 5. The four built environment types are found to be quite distinct of one another. Comparing the cluster centroids to the county-wide statistics (Table 1) shows that the four built environment types align well with the sorting of block groups according to their degree of urbanity.

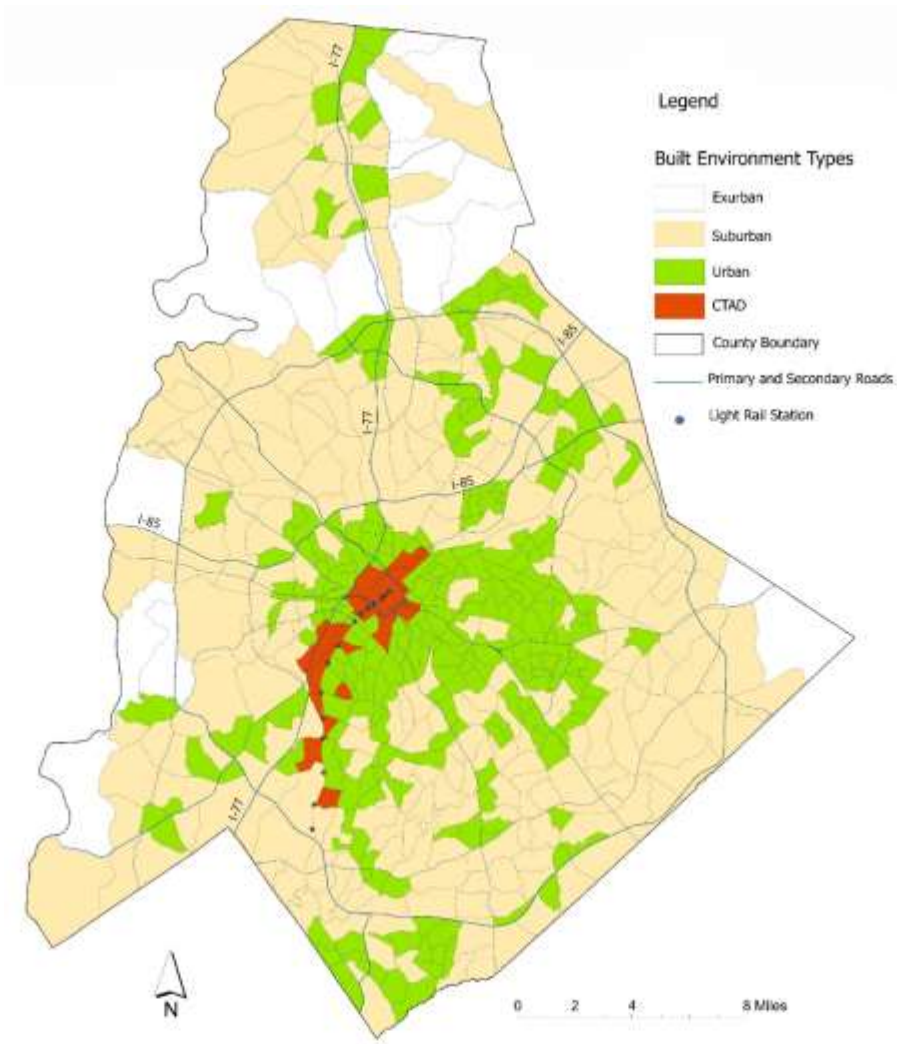


Figure 1. Classification results for built environment types in Mecklenburg County, NC (2015).

Table 5. Built environment type cluster centroids (2015).

Built Environment Related Variable	CTAD	Urban	Suburban	Exurban
Land-use mix (normalized entropy index)	0.0023	0.0016	0.0007	0.0002
Road density (ft/mi^2)	97,172	80,597	52,600	25,239
Intersection density (units/mi^2)	122.95	95.55	52.19	24.43
Housing density (units/mi^2)	4001.7	1875.5	743.2	138.6
Multifamily housing (%)	91.25	28.52	11.95	0
Average size of single-family lots (ft^2)	9590	33,738	30,516	88,570
Jobs-housing ratio	14.691	1.123	1.706	1.010
Jobs accessibility (jobs)	2986	492	851	411

The coefficients for the (non-spatial) Ordinary Least Squares (OLS) model estimated on a sample of 1,092 block groups are reported in Table 6. In this specification, some independent variables were log-transformed to secure homoscedasticity. Based on these results, both CTAD and urban built environment types have statistically significant impact on commuting duration, all else being equal. Specifically, we find that the average commuting duration in CTAD areas was 1.98 minutes shorter than in the reference areas (2.576 – 4.554 = - 1.978) in 2015[1]. However, for urban areas, their 2015 average commuting duration was longer by 0.36 minutes (1.084 – 0.726 = 0.36) compared to reference areas. In addition, the average treatment effects show an increase in commuting duration over time in both CTAD and urban built environment types with coefficients of 2.58 minutes and 1.08 minutes, respectively. These two coefficients are equal to the difference in differences between treatment groups and reference groups over time as shown in equations 2 and 3. In these equations, *y* denotes the commuting duration and *base* denotes the reference group formed of exurban and suburban block groups.

$$CTAD * year\ 2015\ Coefficient = (y_{CTAD\text{-}2015} - y_{CTAD\text{-}2000}) - (y_{base\text{-}2015} - y_{base\text{-}2000}) \quad (2)$$

$$Urban * year\ 2015\ Coefficient = (y_{Urban\text{-}2015} - y_{Urban\text{-}2000}) - (y_{base\text{-}2015} - y_{base\text{-}2000}) \quad (3)$$

In contrast, exurban and suburban built environment types have experienced a reduction in commuting duration of 1.52 minutes from 2000 to 2015. Sociodemographic control variables including white percentage, African American percentage, Asian percentage, median housing value, median age, Ph.D. degree holders, CBD proximity and car ownership all have statistically significant impact on commuting duration, all else being equal.

[1] CTAD areas were not defined in 2000.

Table 6. Non-spatial panel data model results.

Independent Variable	Coefficient	t-Stat	z-Probability
Intercept	28.283	5.072	0.000
Year 2015	−1.519	−3.626	0.000
CTAD	−4.554	−5.159	0.000
CTAD × Year 2015	2.576	2.202	0.028
Urban	−0.726	−1.830	0.068
Urban × Year 2015	1.084	2.191	0.029
White population (%)	−0.096	−4.716	0.000
African American population (%)	−0.065	−3.302	0.001
Log (Asian population (%))	−0.351	−1.915	0.056
Log (Housing density (units/mi^2))	0.045	0.276	0.782
Log (Median household income (dollars))	0.618	1.232	0.218
Log (Median housing value (dollars))	−1.160	−3.327	0.001
Log (Median population age (years))	2.356	2.905	0.004
Associate's degree holder (%)	0.068	1.585	0.113
Bachelor's degree holders (%)	0.000	0.001	0.999
Log (Ph.D. degree holders (%))	−0.596	−2.091	0.037
CBD proximity (mi)	0.709	16.776	0.000
No vehicle available (%)	0.094	4.688	0.000
R^2		0.342	
Adjusted R^2		0.331	
N		1092	

While the OLS results were an important step in the model building process, we now move on to the analysis with a similarly specified spatial Durbin model for which estimation results are reported in Table 7. This model was selected among all the spatial models that were tested, including spatial lag, spatial error and spatial error Durbin models, as the best performing model, while avoiding bias and maintaining efficiency. The best model is selected based on the goodness of fit measures of R^2, squared correlation coefficient between the fitted values and observed values, maximum likelihood, and after investigating the plots and maps of model residuals. The R^2 of the spatial Durbin model is 0.422, which is an improvement of 0.08 over the OLS model.

According to the spatial Durbin model, after controlling for spatial dependence in the dependent variable and in independent variables, CTAD areas have an impact on commuting duration that is significantly different statistically from the suburban and exurban environments, all else being equal. The average commuting duration of CTAD areas was shorter by 0.99 minutes (2.614 - 3.605 = -0.991) compared to other areas in 2015. Similar to the non-spatial model, commuting duration has increased in both CTAD and urban areas with average effects of 2.61 minutes and 0.97 minutes, respectively, between 2000 and 2015 (Table 7). The coefficient of 2.61 for CTAD areas indicates changes over time in differences in commuting duration between CTAD and reference groups (exurban and suburban). In addition, the coefficient of 0.96 for urban areas shows the differences in commuting duration (minutes) between the urban areas and reference groups (exurban and suburban) and between the two years of 2000 and 2015. Mathematically, the rationale for what these coefficients mean is provided in equations 1 and 2, respectively. Exurban and suburban areas, on the other hand, experienced decrease in commuting duration in the order of 1.33 minutes on average.

Like in the non-spatial model, sociodemographic control variables including white percentage, African American percentage, Asian percentage, median housing value, median age, Ph.D. degree holders, CBD proximity, and car ownership have statistically significant impact on commuting duration, all else being equal. Overall, the spatial Durbin model shows results in line with those of the OLS model.

Table 7. Spatial Durbin panel data model results.

Variable	Coefficient	t-Stat	z-Probability
Intercept	−7.735	−0.842	0.400
Year 2015	−1.329	−3.307	0.001
CTAD	−3.605	−4.281	0.000
CTAD × Year 2015	2.614	2.333	0.020
Urban	−0.521	−1.386	0.166
Urban × Year 2015	0.968	2.095	0.036
White population (%)	−0.073	−3.749	0.000
Black population (%)	−0.050	−2.653	0.008
Log (Asian population (%))	−0.307	−1.751	0.080
Log (Housing density (units/mi^2))	−0.057	−0.351	0.726
Log (Median household income (dollars))	0.277	0.595	0.552
Log (Median housing value (dollars))	−0.762	−2.237	0.025
Log (Median population age (years))	1.899	2.442	0.015
Associate's degree holders (%)	0.023	0.568	0.570
Bachelor's degree holders (%)	0.009	0.515	0.606
Log (PhD degree holders (%))	−0.477	−1.769	0.077
CBD proximity (mi)	0.922	11.987	0.000
No vehicle available (%)	0.110	5.366	0.000
Spatial lag	0.327	9.118	0.000
Spatial error terms:			
W×White population (%)	−0.034	−0.839	0.401
W×Black population (%)	−0.022	−0.573	0.566
W×log (Asian population (%))	−0.104	−0.272	0.786
W×log (Housing density (units/mi^2))	0.528	1.986	0.047
W×log (Median household income (dollars))	3.552	14.752	0.000
W×log (Median housing value (dollars))	−1.406	−2.095	0.036
W×log (Median population age (years))	1.428	0.964	0.335
W×Associate's degree holders (%)	0.290	3.318	0.001
W×Bachelor's degree holders (%)	−0.041	−1.348	0.178
W×log (PhD degree holders (%))	0.080	0.137	0.891
W×CBD proximity (mi)	−0.486	−5.306	0.000
W×No vehicle (%)	−0.009	−0.249	0.803
R-squared		0.422	
Squared correlation coefficient		0.388	
Log Likelihood		−2973.560	
N		1092	

2.5. Discussion and conclusions

In order to alleviate the economic, environmental and social consequences of high-cost travel behavior such as long travel distance or duration, and of car dependent travel, new urban planning and design ideas have been advocated, such as new urbanism, sustainable urban development and Transit-Oriented Development. These approaches purport to reduce travel demand and consumption by encouraging density, diversity, design, destination accessibility and reduced distance to transit. There is a considerable body of literature on the relationship between the built environment and travel behavior, which stands as a core element of these innovations. These studies have reached diverse conclusions, with some finding statistically significant reduction of travel and others with very slight or no impact in neighborhoods with denser and more compact development.

We pointed out some methodological issues in this body of literature. One of these methodological issues stems from self-section embedded in the data. Self-selection means that the difference in travel behavior associated with the built environment may have less to do with differences in the built environment itself than with people who in fact self-select themselves into built environments with specific characteristics owing to their travel preferences. Spatial dependence is a second consideration that may distort the true nature of the relationship between built environment and travel behavior components including the commuting duration studied here. To alleviate these issues, this study used longitudinal design as a method for removing unobserved factors; the design also controlled for spatial dependence by accounting for the spatial dependence econometrically. By focusing on the difference in differences, we aimed at sorting the impact of the built environment on travel demand.

This article empirically studied the impact of built environment types on commuting duration as a core travel behavior component. There were two hypotheses. First, built environment types have impact on commuting duration. Second, areas whose urban fabric is deeply associated with density, diversity, design, destination accessibility and reduced distance to transit have shorter commuting duration.

To test the two stated hypotheses, the built environment in 2015 was classified into types of exurban, suburban, urban and CTAD on the basis of built environment factors indicative of the 5 Ds. Their differential impact was then evaluated using spatial panel data models. Spatial econometric results show that the built environment in both CTAD and urban areas has statistically significant impact on commuting duration in comparison to other built environments, namely the city's suburbs and exurbs. Both the CTAD and urban areas had shorter commuting durations. However, our analysis reveals that this impact consists in an increase in average treatment effect over time, that is between 2000 and 2015 in our case study, while exurban and suburban areas have experienced a reduction in commuting duration over this period, even after controlling for spatial dependence. It decreased by 1.33 minutes per commute trip on average in the spatial Durbin model. In conclusion, the findings of this case study confirm that the built environment is not a neutral context within which travel behavior happens to take place. Statistical evidence tells us that commute duration varies across built environment types. Results of our case study are similar to others that found a statistically significant impact, although practically small. They are not consistent with the notion that a built environment developed according to the 5 D principles dampens travel demand, however. Instead, they align with studies that have argued that compact developments lead to higher traffic congestion and greater travel duration. Given the importance of such results for policy making in cities that face the thorny challenges of balancing continued growth and land development, on the one hand, and mobility imperatives, on the other, we find our work contributes to understand the link between travel demand and the built environment. Importantly, we also realize that it does not settle the matter in simple terms, while opening the door to alternative explanations that are broached below.

The apparent lack of consistency between the results of our analysis and large segments of the extant literature begs the question of the possible causes for such differences. In this respect, we believe a highly pertinent observation is that most of the empirical literature treat relatively populous cities, such as Los Angeles, CA, New York Metropolitan Area, San Diego, CA, Boston, MA, and San Francisco Bay Area, CA, that are quite mature. Population densities are higher in these case studies than in Charlotte and transit systems including subway or light rail systems have been operating for much longer periods. More transit options are available in these areas with strong connectivity between different mobility options such as bus and rail. Charlotte stands in sharp contrast as it has an established reputation as a sprawling city (Song and Knapp, 2004, Wilson and Song, 2009) and has only rather recently experienced a process of densification (Delmelle et al., 2014), while the Lynx light rail service is still quite new (it started operation in Fall 2007) and offers low coverage and low connectivity, and has low ridership. These striking differences have several implications on the study of the relationship between the built environment and travel demand. First, by 2015, residential density of Charlotte's light rail corridor had increased. This increase may have exacerbated the traffic congestion in those compact areas as the transit system still provided low coverage and low connectivity, which led to an increase in commuting duration of 2.61 minutes in CTAD and 0.96 minutes in urban areas. Second, in 2015, the transit and population density may not have been in conditions conducive to encouraging people living in CTAD or urban areas to use transit instead of their personal vehicles. Residents of CTAD areas may still find their personal vehicles more convenient. Thus, these land development and transit network properties may have led to traffic congestion and to an increase in commuting duration in compact areas, as policy makers and planners failed to consider their various aspects comprehensively, such as connectivity, coverage, availability of parking facilities near transit stations, and public transit ridership. Along with this point, it can be argued that travel mode is as important as travel duration. As a result, it will be insightful to study commuting duration by travel mode, such as driving to work or using public transit. Third, suburban and exurban areas of Charlotte have themselves evolved tremendously over the study period, with a frantic pace of land development but also with new highway infrastructure that has succeeded in curbing traffic congestion in these areas.

We contend these differences are critical because, in urban planning and design, compact developments and transit-oriented plans are long run in nature. Our analysis spans 15 years (2000 to 2015), yet Charlotte is still in search of a "steady state" in its development form and in its transportation infrastructure. For instance, the Charlotte community has recently approved a new transit vision plan (Charlotte Area Transit System, 2019) that will dramatically enhance the current single-line Lynx service to a fully built-out system, along with integrated land-use planning and transit-oriented development as its cornerstones. As evidence of the evolving mindset and priorities of the community, the City of Charlotte is in the midst of the approval and implementation of a new Comprehensive Plan (City of Charlotte, 2020) that espouses the principles of community-centered development with "10-minute neighborhoods" as its first goal. Hence, 2015 shows an urban landscape full of nuances that cannot be fully assessed outside of the long-term adjustments towards a new land-use-mobility "equilibrium". It remains to be seen whether impacts may change in future years, when the transit network properties are improved and population density is higher. In future conditions of a more populous city, people may have a greater tendency towards using transit.

In addition, the majority of the research on the impact of the built environment on reducing travel consumption has been done on all trip types, while this work studied work travel. This is important to consider because the response of non-work trips might well be different from work trips in CTAD and urban areas. The 5 Ds may have a greater impact on non-work travel duration or distance than on work travel since the options for services for daily needs and their accessibility increases in compact and diverse neighborhoods.

Given the points raised on the status of residential development, transport infrastructure and mobility in Mecklenburg county in 2015 and given the broader implication this may have on settling the nature of the relationship between the built environment and travel consumption, an agenda can be laid out for future research. First, we propose that cities with diverse degrees of development and transit properties should be studies through comparative approaches to gain insights and shed a better light on the important considerations discussed above.

Second, the longitudinal design method for alleviating the selection bias is based on the assumption that travel preferences are constant over time, while this may not hold true in fast changing urban context, such as with the emergence of new housing options and new mobility options (e.g., novel Mobility-as-a-Service (MaaS) providers). Having access to primary survey data on travel preferences may help in better understanding of the role of selection bias for future work.

In addition, given the potential changes in traffic conditions in compact areas during a period of fast transformation of the urban fabric, it would be helpful to control for traffic congestion effects brought about by localized population and employment growth. Also, with the growing multimodality of urban travel, some measures on walk-and-ride, park-and-ride, and other first/last mile coping behaviors should be helpful to the modeling work.

Lastly, the relationship between the built environment and travel consumption is a multifaceted and multidirectional phenomenon. Bidirectional relationships are possible between different components of travel behavior such as travel duration, travel distance, travel mode, trip type, trip frequency and mode choice, different factors of the built environment and travel preferences. As a result, more holistic models that can investigate complex relationships in a multidirectional phenomenon such as structural equation modeling should be considered for future work.

CHAPTER 3: NON-BUSINESS SERVICES PERFORMANCE FORECASTING FOR SMALL AREA USING A SPATIOTEMPORAL DEEP LEARNING METHOD

Abstract

Having information about the future of economic outcomes in advance is critical in policy making and decision making for both the public and private sectors. However, the urban economics literature mainly focuses on causal relationships and has paid lip service to forecasting economic dynamics. In this research, we study the performance of non-business services, including retail, accommodation and food services, and other services, and predict their outcomes for three years ahead at fine spatial resolution. Given the complexity of relationships between social, environmental, and economic factors, a multivariate forecasting framework is used to forecast three economic indicators of employment, business sales volume, and labor productivity. A number of covariates measuring business performance, sociodemographic characteristics, agglomeration economies forces, built environment, and real estate investments are incorporated in the multivariate model. The spatiotemporal forecasting model is developed using Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNN) to predict economic outcomes. The model is trained and tested in Mecklenburg County, NC, at the resolution of census block groups. The predictive performance of the models is judged to be high for all three outcome measures using the Mean Absolute Error (MAE) metric. This is all the more remarkable given the notorious difficulties to do small area economic forecasting.

Keywords

Non-business services, time series forecasting, deep learning, Recurrent Neural Networks, Long Short-Term Memory, dimensionality reduction

3.1. Introduction

With cities producing 80% of the global gross domestic product (World Bank, 2020), the study of urban economic growth and development is of great importance to maintain and guide the economic and social vitality of many developed and developing nations. In addition, accelerating changes in urban dynamics and decentralization of housing and employment in the form of evolutions from monocentric to polycentric cities have led to fundamental shifts in the spatial distribution of economic activities (Kloosterman & Musterd, 2001). These changes and new dynamics make the study of urban economics increasingly more important. In addition, urban economics has evolved from earlier exogenous growth models to endogenous growth models in which human capital and technology, as the most pivotal growth factors, are internal as opposed to external drivers. According to endogenous growth theories, these factors are the main determinants of sustained growth in the long term (Martin & Sunley, 1998). The urban economics literature covers the intricate interactions between its different elements, such as land use, transportation, land value, housing affordability, market dynamics, and location decisions of households and firms.

However, the majority of research on urban economics has focused on the causal connections between its elements, while only a small fraction have been devoted to predictions and future forecasts. Understanding the variabilities in economic growth and development over time and space is of tremendous importance. Indeed, predictions of economic growth can be used for the identification of emerging subcenters that can change the spatial structure of cities. Identification of these areas is very valuable for real estate investors, developers, and business owners to find locations with better return opportunities in the future. In one study, Berg (2014) analyzed entrepreneurs' location decisions through interviews and found that they usually had limited information on these locations and their future returns. According to his study, most of the location decisions were made by chance. This lack of information provides evidence for the significance of having knowledge about future dynamics beforehand. On the other hand, future forecasts of economic growth can help to identify declining neighborhoods. Being able to anticipate neighborhoods that may be declining in the future can be very helpful for planning, policy making, and decision making in relation to transportation, housing and infrastructure, allocation of resources and services, jobs availability, etc., through the lens of spatial and social equity. In conclusion, future forecasting of economic growth sheds valuable light on future opportunities and threats, although such endeavors suffer from a lack of attention in the urban economics literature. In this study, we develop a spatiotemporal model to forecast the business performance of neighborhoods for three years ahead. We propose a deep learning model, and train and test it using historical data for the years 2002 to 2018.

Urban economics has been largely explained through agglomeration economies or economies of scale. The economies of scale mean that firms and businesses agglomerate and cluster together in the urban space as they gain benefits from this agglomeration depending on their industries and input-output linkages between them and other industries. This scaling effect is a result of various forces such as input-output linkages between industries, common access to the same pool of labor between firms in the same industry, knowledge spillover effects between firms in similar or different industries in close proximity, natural advantage including factors such as mutual access to transportation infrastructure, service providers and consumers (Puga, 2010). Having access to common resources between firms or businesses through the mentioned forces reduces costs such as costs of transportation, labor, inputs such as natural resources, and warehousing and distribution. The discussed agglomeration forces are expected to play an important role as predictors of economic outcomes.

It is noteworthy that agglomeration forces and their benefits vary significantly over industries (Giuliano et al., 2019; Hanlon & Miscio, 2014). A great deal of these variations occurs between industries, whether they are in the group of business services or the group of non-business services. The biggest difference between these two groups is their demand side. Business services provide services to other businesses, while non-business services provide goods and services for individuals as customers. As a result, their input-output linkages, labor force skill level, sales volume, production scale, sensitivity to changes in other industries, and many other factors are different. Given these fundamental differences, it is important to study these industries in separate groups.

Although the US has been experiencing upward trends in the growth of business services in the economy as a result of tremendous advances in technology, non-business services still make up a considerable share of employment (56.57% of total employment in the US in 2017) (US Census Bureau, 2017). Also, they comprise 54.6% of total firms in the US. Additionally, non-business services provide jobs to low-skilled people with lower wages (only 0.38% of the total US payroll while covering 56.57% of total employees in 2017) (US Census Bureau, 2017). Given the accelerating growth in business services and high-skilled jobs, it becomes more important to pay attention to jobs for low-skilled individuals, particularly in growing cities such as Charlotte, NC.

As mentioned earlier, factors affecting non-business services are different from business services. Non-business services depend on foot traffic. As a result, accessibility to these services plays an important role in their growth. The accessibility is for both customers as service consumers and workers as service providers. Higher accessibility for workers improves labor pooling and matching as one of the significant agglomeration forces (Chatman & Noland, 2011). Therefore, it is expected that areas with better transportation accessibility provide more opportunities for the prosperity of non-business services. These accessibility factors include public transit such as light rail, parking availability, the walkability of the area, and accessibility by different modes of transit. Their demand side can be measured by its respective housing market, housing density, trade area, and spending power of neighborhoods in their service area, such as median income. Their supply side also plays an important role, such as accessibility to suppliers and warehousing.

Most studies in urban economics, whether in causal relationships or future predictions, are conducted at large scales, exploring the economic mechanisms in sets of cities or regions (Billings & Johnson, 2016). However, having an understanding of economic mechanisms at the neighborhood level is very important as it can help the location decision-making of firms and businesses. In addition, as explained through one of the well-known problems in geographic analysis, namely the modifiable areal unit problem (Openshaw, 1979), scale of study is an important determinant in analysis outcomes. Also, as building blocks of the regional economy, there is considerable heterogeneity in neighborhood economic and demographic characteristics that need to be studied. Furthermore, even for regional economic strategies, interconnections between neighborhoods and regions, in addition to the heterogeneity in neighborhoods, need to be considered (Weissbourd, 2006). This point becomes more significant in policy making as different interventions and policies are required for different geographic granularities. In addition, recent advancements in data availability in the fine spatiotemporal resolutions of neighborhoods annually make the analysis of economic dynamics now feasible.

This study aims to predict future economic outcomes of non-business services at the census block group level representing neighborhoods in Mecklenburg County, NC, in the United States. Neighborhood economic outcomes are measured by multiple indicators, including employment, business sales volume, and labor productivity as an economic indicator derived from the first two variables.

Given the known interdependencies of sociodemographic, economic, and built environment factors in urban economies, multivariate forecasting is used to predict the neighborhood's productivity. In this predictive model, potentially related covariates such as built environment characteristics, agglomeration forces, business characteristics of the neighborhoods, and real estate investments are taken into account as covariates.

Built environment characteristics have been hypothesized to be associated with the economic development of cities at different geographic levels, such as neighborhoods, particularly the economic development of non-business services such as retail. A range of urban planning and design theoretical approaches have been developed to enhance the economic vitality of neighborhoods, such as new urbanism, mixed-use development, Transit-Oriented Development (TOD), and smart growth (Day, 2003; Hess & Lombardi, 2004; Krueger & Gibbs, 2008). In these approaches, built environment characteristics such as transit proximity, Central business district (CBD) proximity, density, jobs-housing ratio, and land-use mix have been promoted for their positive impact on economic development.

Given these studied impacts, the built environment characteristics are incorporated in the predictive model of this study as potential predictors.

Additionally, attributes of other businesses such as their sales volume, count of employees, and business type (e.g. headquarter, branch, single location, etc.,) are incorporated in the model to capture agglomeration forces such as specialization, urbanization, and location quotient. Furthermore, trade area demographics such as income, race, and age are incorporated into the model as they are expected to have an impact on the neighborhood's economic performance (Kumar & Karande, 2000). Lastly, real estate investments have been known as leading indicators of the economy (Manuele, 2009; Strauss, 2013). An increase in the number of new residential construction or commercial developments can inform us about economic development in advance. Therefore, building permits of different types, such as new residential and commercial constructions and their renovations, are used as leading predictors.

Annual historical data of the years 2002 to 2018 are used for the economic outcomes and their covariates. Given the complexity of relationships between economic predictors and urban dynamics components, and given the limitations in assumptions of traditional econometrics such as linearity, deep learning (DL) is used for the time series forecasting as it is capable of handling the anticipated complexities of the empirical problem under study. Due to the high performance of machine learning (ML) and DL models in predicting socioeconomic dynamics, they have been more widely used in recent years (Arribas-Bel et al., 2021; Delmelle and Nilsson, 2021; Glaeser et al., 2018; Hatami et al., 2023; Rajesh et al., 2021; Reades et al., 2019; Ron-Ferguson et al., 2021), instead of conventional econometric models. In socioeconomic phenomena, there are complex relationships between a large number of features. DL models digest these relationships in high-dimensionality data very well and with few or no limitation. In contrast, in econometric models, assumptions have to be met and data engineering needs to be done carefully.

3.2. Literature review

There is a large body of literature on urban economics. The literature investigates this multifaceted topic from different dimensions, such as land-use/transportation integration, neighborhood change, and sustainability. Most of these studies explore the associations and causal relationships between various aspects of urban dynamics rather than forecasting over a certain planning horizon. Examples include studies investigating the impact of transportation on property values (John, 1996; Martínez & Viegas, 2009; Mohammad et al., 2013; Ryan, 1999; Schmidt et al., 2022), the impact of transportation on retail commercial development (Credit, 2018; Yarbrough, 2014), identifying the determinants of polycentric structures (Liu et al., 2011; Lv et al., 2021), and relationships between economic development and environmental issues (Shafik, 1994). Less attention has been paid to future trends predictions.

3.2.1. Artificial intelligence and data-driven methodologies in studying urban economies

In addition, the growth in high-resolution data availability and recent advancements in data-oriented approaches that originated in computer science, such as machine learning and deep learning, are expected to bring considerable advancements in studies of urban economies and economic geography. However, it is only recently that serious attention has been paid to the applications of artificial intelligence (AI), ML, and DL to problems in these domains of application. Delmelle (2022) conducted a review on recent applications of data-driven and ML methods for predictive and explanatory analysis on neighborhood change processes. Hatami et al. (2023) utilized ML to predict the commuting modal split in Mecklenburg county, NC. They further applied the SHAP ML method to investigate the impact of the built environment on commuting mode share. The ML methods used in this paper resolve complexities such as non-linear associations that conventional econometric models are not able to capture. Delmelle and Nilsson (2021) performed text analysis on real estate advertisements using ML to explore the relationships between advertisements text data and neighborhood socioeconomic characteristics. Their trained ML model performed well in predicting the neighborhood type. This finding indicates potentials of using AI and text analytics for predicting future neighborhood changes before it happens. Arribas-Bel et al. (2021) used the Approximate DBSCAN method to delineate urban areas in Spain using data from 1 million residential and non-residential buildings. Then, they compared their delineated boundaries with the commuting-based boundaries and administrative boundaries and found more consistencies with the commuting-based boundaries. Emphasizing the important role of historical data in some research areas of urban economies such as agricultural productivity and urbanization; urban growth, city structure and transportation; communication, transportation and trade; and migration and interregional mobility, Combes et al. (2021) studied how machine learning methods including the random forest and neural networks can be applied for the treatment of historical data. Kourtit (2021) studied global cities' safety and security, and economic development potentials, namely magnetism. They used variables related to the stated dimensions. Using an advanced sequential cluster dynamics analysis method, they identified clusters of cities. They found that cities' safety and security were important predictors of their economic performance. Ron-Ferguson et al. (2021) used random forest machine learning technique to study urban development and change through demolition and new construction activities. Applying ML, Reades et al. (2019) studied neighborhood change from 2001 to 2011 to predict gentrification in London neighborhoods. Overall, most of the applications of AI in economic geography are very recent and not much attention has been paid to this until very recently.

3.2.2. Fine spatial granularity in urban economics studies

In the literature on urban economies, studies on future forecasts have usually been on predictions of population (Booth, 2006) in large scale and coarse resolutions or urban growth and land use change using satellite imagery datasets (Aburas, 2016; Allen & Lu, 2003; Batty & Torrens, 2001; Clarke & Gaydos, 1998; Park et al., 2011; Triantakonstantis & Mountrakis, 2012), not socioeconomic data. Only a small number of studies have investigated forecasts and predictions of economic growth with actual economic indicators using socioeconomic data. Pentland (2020) studied the interactions between communities and flows of ideas between them as an important predictor of economic growth outcomes. He found that diversity of ideas and interactions between agents at individual, business and national levels played an important role in predicting economic growth. Chong et al. (2020) predicted economic prosperity and growth at the neighborhood level in Istanbul, Turkey, Beijing, China, and several cities in the United States. They used the diversity of consumed goods and services within neighborhoods and interactions and flows between them as predictors of economic growth and prosperity. Wei (2018) predicted economic growth in the presence and absence of high-speed rail in Changsha City, China, to evaluate the impact of high-speed rail on economic growth. Applying regression and Gray prediction models, they used data from 2001 to 2009 and predicted the economic growth from 2010 to 2015. Using data from remote sensing and interviews, Warth et al. (2020) predicted the socioeconomic indicators, including income, building assets, and educational attainment, to support future infrastructure planning.

3.2.3. Economic heterogeneity across industries

Despite the presence of heterogeneity over industries in urban economics, a majority of the forecasting literature, including the ones discussed above, has been on economic growth in general and not on specific industries like non-business services. Non-business services such as retail, restaurant and other services to non-business clients have been known as important drivers of economic growth or decline at the neighborhood level in the field of community development (Koebel, 2014; Zukin et al., 2009). The contribution of these services to the neighborhood and city economy has been one of the main focal areas of newer theoretical planning approaches such as new urbanism, Transit-Oriented Development, and smart growth. A number of studies have been dedicated to the relationships between the growth of these industries and neighborhood economic development. Waxman (1999) studied the revitalization of inner-city neighborhood business districts through economic theories of retail and commercial revitalization in Dorchester, Massachusetts. Similarly, Koebel (2014) studied neighborhood revitalization through commercial revitalization and growth in the number of retail stores and service establishments. This study argued that most of the research focus has been on the impacts of large-scale commercial developments, while overlooking the importance of disadvantaged neighborhood commercial districts and their influence on the vitality of neighborhoods. The study stated that growth or decline in the number of retail stores and service establishments was a function of market factors, including trade area demographics and demand characteristics, and non-market factors, including discrimination as a result of race and ethnicity. Dolega and Lord (2020) investigated the economic growth and decline of neighborhoods by dynamic trajectories of retail business performances between the years 2014 to 2016 in the Liverpool city region, UK. They found that the trends of retail economic growth varied over space in this region. Focusing on retail grocery, Pothukuchi (2005) investigated initiatives for increasing grocery investment and found variations in those initiatives in different sites. They found that systematic initiatives for the entire city were not common and recommended that planners increase investments in retail groceries in disadvantaged areas to promote economic development.

3.2.4. Forecasting non-business services

Recently, more attention has been paid to future forecasts of non-business services in urban economies. Using Yelp social media data, (Glaeser et al., 2018) studied the relationship between the number of restaurants, bars, cafes, and grocery stores, and gentrification in New York City zip codes and found a significant association. They found that local businesses could predict an increase in housing prices and gentrification as leading indicators. Rajesh et al. (2021) measured the resiliency of the retail industry of urban centers using five resilience indicators of the vacancy rate, retail turnover dynamics, growing sales history tenant structure, and economic crises from 2010 to 2019. On this basis, the study predicted the resiliency of the retail industry in urban centers in 2020 using grey prediction and moving probability-based Markov models.

3.2. Research Design, Methods and Data

3.2.1. Forecasting framework

In this study, we develop a framework to forecast the performance of non-business services at a fine spatial resolution (Figure 2). Economic performance is measured by several indicators that complement each other, each one capturing a specific perspective on economic performance locally. The first economic indicator is employment. Employment trends in non-business services are essential to investigate as these businesses mainly employ low-skilled labor. In developed countries, high-skill jobs such as business service activities are increasingly taking a more important part in the economy. Therefore, as the low skilled workforce is at risk of being cut back and displaced by technology, it is essential to pay attention to employment in businesses with low-skilled labor. The second indicator is the business sales volume. This variable is selected as sales of non-business services such as retailing indicate the spending power of consumers and demands. Finally, the last economic indicator is the average sales volume per employee, which tracks businesses' productivity. Generated using the first two variables, the latter indicator is of great importance as it shows the profitability of businesses. Knowing about the future aggregate profitability of a neighborhood is critical from the businesses' perspective as it helps them to locate in or relocate to areas with higher returns.

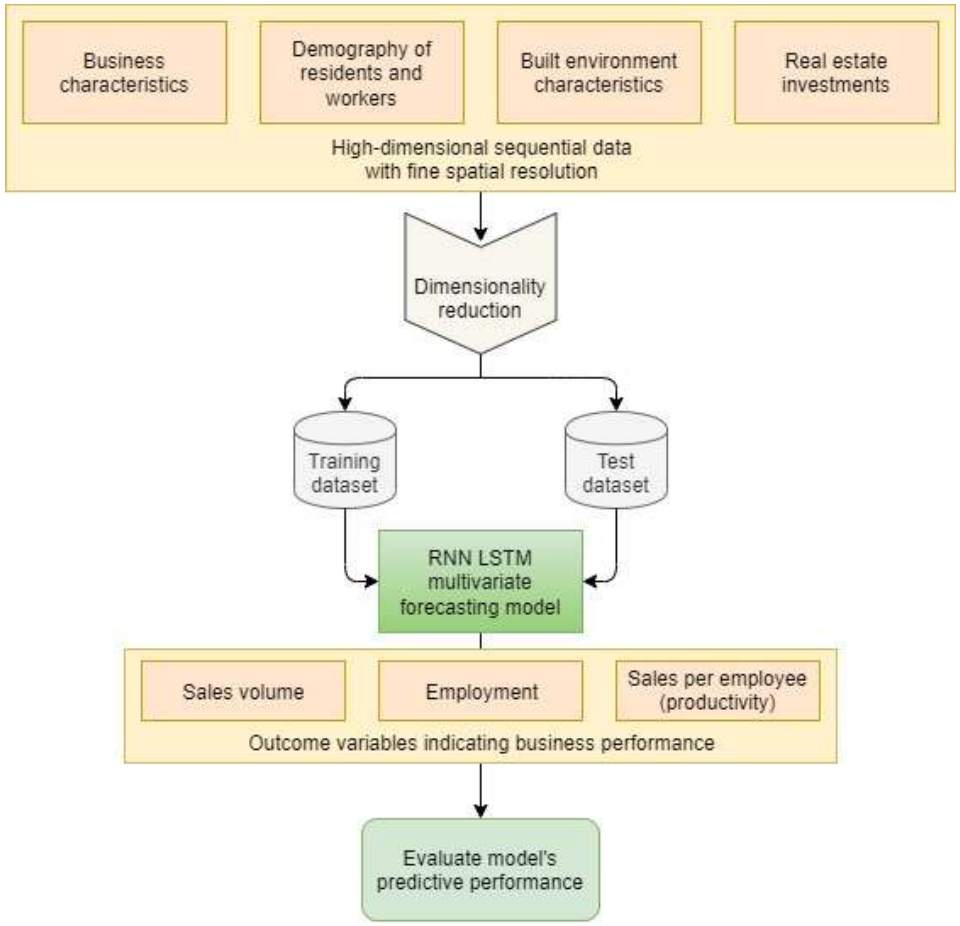


Figure 2. Proposed framework for forecasting business performance of neighborhoods using an RNN LSTM multivariate model.

As a result of the complexity of socioeconomic phenomena at fine spatial resolution and of the interdependencies between covariates, multivariate forecasting is used. The high dimensionality of covariates is collapsed into classes of business characteristics, demography of residents and workers, built environment characteristics, and real estate investments. Due to the forces of agglomeration economies, characteristics of businesses within and outside of the studied industry are expected to have an impact on business performance. The demographic characteristics of residents and workers are expected to be strong predictors as they show the characteristics of consumers and of the labor pool, which link up to the demand and supply of the activities being forecasted. Since non-business services are dependent on foot traffic, built environment factors such as accessibility play an important role in the neighborhood's business performance. Lastly, real estate investments are used as a leading indicator of the economy.

Due to the complexity of the subject, our framework proposes to use DL methods to identify the combination of predictors that most effectively forecasts neighborhood economic performance. DL methods are capable of digesting complex situations in which there are non-linear relationships between covariates, when there is a large number of interdependent features, the amount of data is large, or when there are multiple seasonality and trends in the time series data. Another benefit of using the DL method over traditional econometrics and machine learning is that it requires less domain expertise, making it useful across a range of disciplines. In addition, agent-based modeling is another common method in urban dynamics forecasting. However, there are some limitations in these models and other simulation-based models. All of them have a rigid set of assumptions that simplify the phenomenon. They are extremely reliant on and sensitive to the model's input variables. The prediction models will also be univariate rather than multivariate. In contrast, cutting-edge data-driven methodologies, such as DL, draw knowledge from various aspects of the phenomena captured in the data. For the purposes of neighborhood economic performance forecasting, we propose a multi-output multistep multivariate time series forecasting approach that is based on the Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNN).

As an extension of artificial neural networks, deep neural networks are highly capable of capturing complexities due to having a large number of hidden layers in their architecture. An artificial neural network is composed of an input layer, an output layer, and a hidden layer(s). Neurons in the mentioned layers are connected through some links associated with weights and biases. Through the learning process, the model finds the optimized values of these parameters by backpropagation. A deeper network with a higher number of hidden layers can capture more features and complexities in data (Schmidhuber, 2015). As a result of the mentioned power of deep neural networks, along with the availability of high-performance computers, these models have become widespread in a variety of application areas, including in urban economics (Grekousis, 2019; He et al., 2018; Ma et al., 2023; Zheng et al., 2023). Furthermore, as a subset of DL methods, recurrent neural networks can digest complex relationships in sequential data. In recurrent neural networks, the output of the hidden layer at each time is saved and fed back to the layer while keeping a hidden state. Figure 3 shows a simple illustration of the recurrent neural network architecture. U, W, and V are model parameters that are optimized through the training process. As shown in this figure and in equation 1,

$$h_t = F_c(h_{t-1}, x_t) \tag{1}$$

the hidden layer receives information from both the input layer and earlier stages of itself. h_t is the new state; f_c is the function with parameter c; h_{t-1} is the old state; and x_t is the input vector at time step t. Given the high performance of recurrent neural networks in analyzing sequential data, they have received great attention in the literature of different research areas, including urban analytics (Khusni et al., 2020; Nikparvar et al., 2021; Pan et al., 2022; Zhene et al., 2018). Moreover, the LSTM model enhances the model's performance since it resolves the vanishing/exploding gradient issue in recurrent neural networks. The LSTM model learns long-term dependencies and resolves the gradient-related issues through the cell state. The cell state is regulated by gates for removing or adding information (Graves and Graves, 2012).

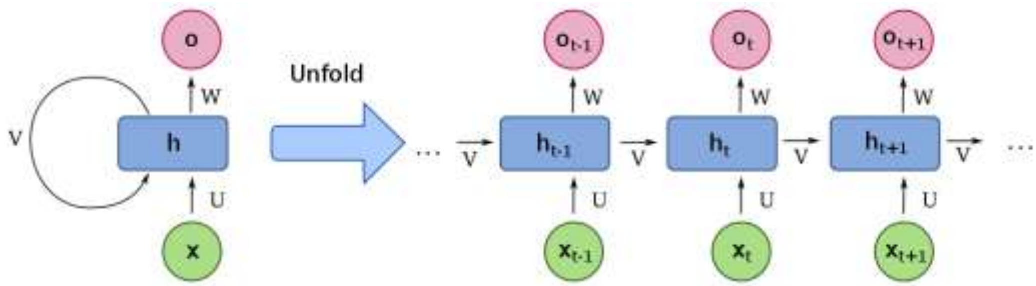


Figure 3. Simple illustration of recurrent neural network

Before training the forecasting model, since we have a large number of features, we perform dimensionality reduction. After performing the dimensionality reduction, data is split into training and test sets, with 80% and 20% proportions respectively, to fit the proposed model and evaluate its performance. The model's predictive performance is evaluated using the mean absolute error metric. The model is developed in Python using the Keras library (Chollet, 2015).

3.2.2. Study area

As a case study, we use Mecklenburg County, NC, and forecasts are developed for geographic neighborhoods, which are taken to be census block groups. As the most populous county in NC until 2020 and with the city of Charlotte as its seat, Mecklenburg County has been experiencing fast population growth and accelerated economic development over the past decades, which makes it a well suited area of study.

The county's population increased 58%, from 700,458 in 2000 (U.S. Census Bureau; 2021a) to an estimated population of 1,110,356 in 2019 (U.S. Census Bureau; 2021b). Total employment has increased by 42% from 2002 to 2018 (LEHD, 2021).

The three industries of Retail Trade, Accommodation and Food Services, and Other Services studied in the second part of this study comprised 21.29% of the total employment in Mecklenburg county in 2018, with the Retail Trade and Accommodation and Food Services in the top 5 industries among all 2-digit NAICS codes in terms of share of employment. These three industries had significant growth in Mecklenburg county from 2002 to 2018, with percentage changes of 35%, 85%, and 33%, respectively (LEHD, 2021).

Anticipating the continued growth of population and of the local economy, some plans were developed for this county to better integrate land use, urban design, and transportation systems, such as the 2025 Transit/Land-Use Plan, which was finalized in 1992. This plan entails five transit corridors and developments along them; the first corridor started operating in 2007 as the Blue Line of the Lynx light rail service. The establishment of this light rail line has led to developments incorporating high density, high diversity, low distance to transit, high destination accessibility, and careful design (known as the 5D design elements), including transit-oriented development projects in proximity to the rail transit stations. These developments place a compact mixture of residential, office, retail, and consumer services along the transit corridors. In this study, the annual historical data for the years between 2002 to 2018 are used to forecast the future years' economic outcomes and test the model's predictive performance.

3.2.3. Data sources

Data on sales volume and employment in the three studied non-business services, which are used as indicators of the neighborhoods' business performance, are collected from the US Businesses dataset from the data source Data Axle (Reference Solutions)

(Data Axle, 2021). Sales volume is adjusted for inflation in the data source. Employment and sales volume data are reported by year from 2000 to 2018 for each business. Then, data is aggregated to the census block group boundaries for each industry based on the reported NAICS code of the business and for each year. Also, the count of businesses by industry is calculated for each census block group and year from Data Axle's US Businesses dataset.

In addition, data for the count of employees by industry, residents' and workers' income, age, educational attainment, race, or ethnicity are sourced from the Longitudinal Employer-Household Dynamics (U.S. Census Bureau, 2021c) dataset. These data are available in the census blocks for each year between 2002 and 2018. The census block data are aggregated to the census block group boundaries for each year. Data for capturing real estate investments are collected from the building permits dataset Basic Report for Posse Permits from the Mecklenburg County Open Data (2021). This dataset includes all issued building permits in Mecklenburg county for different categories, such as new residential constructions, new commercial constructions, residential and commercial renovations, and demolitions. Since real estate investments and building permits are considered leading indicators of the economy, the building permits data are collected with a one-year lag (Strauss, 2013). In other words, the study period in this research is 2002 to 2018, and the building permits are incorporated in the model for the years 2001 to 2017. Furthermore, variables representing the built environment, such as proximity to rail stations or proximity to the Charlotte CBD, are calculated from GIS shapefile data collected from Mecklenburg County's GIS Center (2021).

3.2.4. Data description and processing

In this study, the economic performance of neighborhoods is forecasted for specific non-business services for future years. These industries include retailing, accommodation and food services, and other services with 2-digit North American Industry Classification System (NAICS) codes of 44-45, 72, and 81. These three industries are selected due to their close similarities from different aspects. They all provide services to individuals and are dependent on population foot traffic. They often operate with low-skilled labor. They are dependent on the spending power of residents in proximity to the points of sale and service delivery. Due to their dependence on foot traffic, accessibility measures such as transit proximity and CBD proximity are expected to have an impact on their performance. The historical data from 2002 to 2018 are used with an annual temporal resolution. The spatial resolution is that of census block groups representing neighborhoods.

As discussed in the forecasting framework section, three outcome variables are developed and used to predict the business performance of neighborhoods. Descriptive statistics of these three indicators are shown in table 8.

Table 8. Descriptive statistics of outcome variables (observed values in all years and all block groups)

Outcome variable	Count	Mean	Standard Deviation	Min	Median	Max
Employment (person)	9,435	223.31	430.32	0	75	533
Sales volume ($1,000)	9,435	38,000.21	74,805.64	0	10,894.33	718,762
Sales per employee ($1,000/person)	9,435	119.01	74.32	0	107.04	743.44

Covariates that cover forces of agglomeration economies, including colocation using the number of businesses in the same industry, number of employed residents of each census block group indicating the labor pooling, and location quotient or localization is used. The location quotient is an index of a block group’s specialization in non-business services relative to that of the entire county. It is calculated as the ratio of non-business services to all businesses in a block group divided by the same ratio in the entire county. In addition, characteristics of businesses, including the average firm's size and average firm's age, are included. Built environment-related variables, including CBD proximity, proximity to light rail, and jobs-housing ratio, are incorporated as other potential predictors. CBD proximity of each block group is calculated as the distance between the CBD and the block group. Proximity to light rail is taken as the distance between each block group and its closest transit station following a straight line. Sociodemographic variables, including race, education attainment, the mean per capita income of workers and residents, and the mean age of workers and residents, are added as control variables. Lastly, a binary variable indicating the economic recession in 2009 and 2010 is added to the forecasting model and serves as a fixed effect on neighborhood performance.

In this study, we have more than 300 features. As a result, we used principal components analysis (PCA) as a powerful dimensionality reduction technique. We set the number of components to four as they collectively explain 98% of the cumulative variance. The PCA reduces the runtime tremendously as our dataset has more than 300 features. It also improved the model performance significantly. After the dimensionality reduction step, data is split into training-validation and testing sets with 80%-20% shares, respectively. We predict the last three years of data as 2016, 2017, and 2018 as the 20% split to test the model performance. Summary statistics of outcome variables for the three forecasting years are shown in table 9.

Table 9. Descriptive statistics of outcome variables (2016, 2017, 2018 average observed values)

Outcome variable	Year	Count	Mean	std	Min	Median	Max
Employment (count)	Three-year average	1,665	262.51	481.19	0	98.67	5,333
	2016	555	266.03	515.64	0	84.67	5,305.33
	2017	555	274.89	528.27	0	90.33	5,333
	2018	555	246.61	387.45	0	116	3531.67
Sales volume ($1000)	Three-year average	1,665	42,643.60	81,918.40	0	12,391	701,299
	2016	555	47,051.80	90,148.60	0	13,599	701,299
	2017	555	47,509.40	90,011.50	0	14,024	683,187
	2018	555	33,369.50	61,627.10	110.67	10,471	513,898
Sales per employee ($1000/employee count)	Three-year average	1,665	126.39	82.91	0	112.21	600
	2016	555	135.94	86.92	0	123.68	588.19
	2017	555	138.75	89.83	0	125.38	600
	2018	555	104.50	65.54	4.57	94.74	403.89

To improve the model performance, we employed a data augmentation technique by training the model on data of all block groups together, instead of training and predicting each block group individually. We used a window size of 6 years for the training part. The window size of 6 is chosen based on our investigations of temporal dependence between the data. Data is highly correlated for up to 6 years prior. After training the LSTM-RNN model, forecasting is done for the next 3 years for each census block group individually. Out-of-sample validation is done using the remaining 20% of the data. The mean absolute error value is used for model performance evaluation. Lastly, in the first set of results, we had a small number (less than 10) of negative predictions close to zero for employment and sales volume variables, which is a problem commonly encountered in time series forecasting (Burkom et al., 2007). In order to avoid negative predictions, as counts cannot be negative, the logarithm of the outcome variables was taken before training the model, and then, predictions were made using the exponential function. Furthermore, the spatial lag effect was tested in the models by aggregating the values of neighboring units for all variables related to the residential area characteristics (RAC data from the LEHD dataset). Results did not show improvements in fit in comparison with the non-spatial models and are therefore not presented in detail hereunder. One potential reason is the fine spatial resolution of the study. Block groups share similarities with their neighboring units due to their small size in relation to the spatial scope of the study.

3.3. Results

In this study, three economic indicators of local economic performance that involve employment, business sales volume, and labor productivity, respectively, are forecasted in Mecklenburg county census block groups for the years 2016, 2017, and 2018.

The model predictive performance is evaluated for all three outcome variables by the mean absolute error (MAE). The MAE is reported as an average over the three forecast years, and for each forecast year individually. Given the variable range of observed values across block groups, the average MAE is normalized by both the mean and range of observed values in order to better understand the errors in comparison with observed values. The normalized MAE metrics are used as they provide us with information about the magnitude of the error relative to the observed values. The normalized MAE values are more informative as we have a wide range of observed values across all block groups. In addition, this metric is used instead of percentage error, which is one of the most common relative error measures, since the practical magnitude of the percentage error depends on the range of observed values as well. Therefore, normalized error metrics are used to evaluate the predictive performance of the models. As indicated in table 10, for all three outcome variables, the MAE values normalized by range are less than 0.04 for each of the three predictive years. This shows the high predictive performance of the proposed model. The MAEs normalized by mean have values of 0.23, 0.21, and 0.16 for employment, sales volume, and labor productivity, respectively. Taken together, these values show that our model is fully capable of predicting the neighborhood business performance with a high accuracy, that is 97% and higher when compared with the range values, and an accuracy of 76% and higher when compared with the mean values. In addition, the table shows the variability of the performance metrics over the years. The model performs better for the first and second years, in comparison with the third year, as is typical of forecasting over an extended planning horizon.

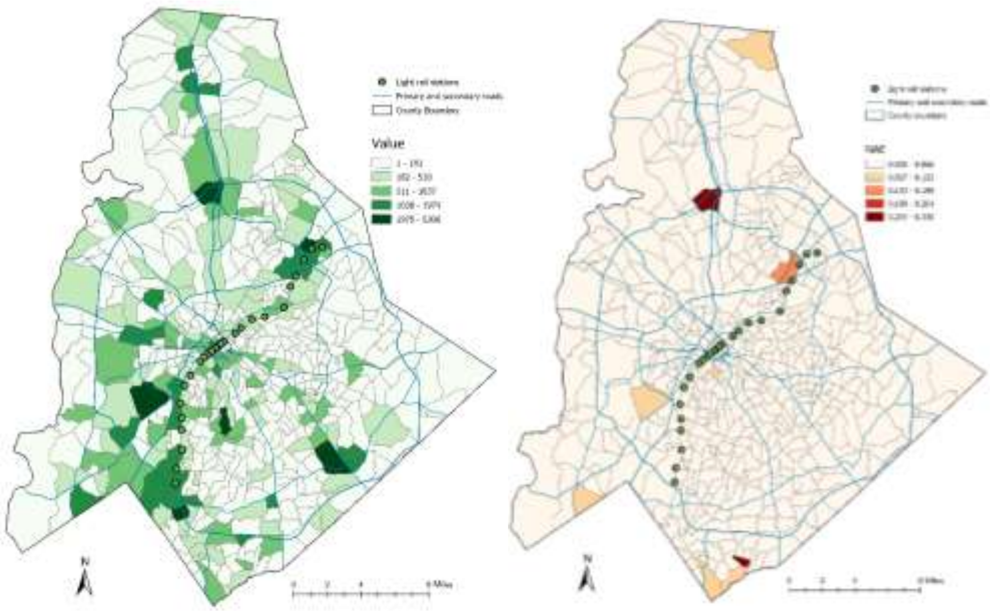
Table 10. Forecasting model performance evaluation using Mean Absolute Error (MAE) and Normalized Mean Absolute Error (NMAE) (normalized by the observed mean and by the observed range), by forecast year.

Evaluation metric	**Forecast year**	**Predicted variable**		
		Employment	Sales volume	Average sales per employee
MAE	All years	61.846	8782.548	20.69
NMAE (by mean)	All years	0.235	0.206	0.164
NMAE (by range)	All years	0.011	0.012	0.034
MAE	2016	35.983	8782.548	14.544
MAE	2017	37.998	7370.391	13.341
MAE	2018	111.558	11956.066	34.185
NMAE (by mean)	2016	0.135	0.187	0.107
NMAE (by mean)	2017	0.138	0.155	0.096
NMAE (by mean)	2018	0.452	0.358	0.327
NMAE (by range)	2016	0.007	0.012	0.025
NMAE (by range)	2017	0.007	0.011	0.022
NMAE (by range)	2018	0.031	0.023	0.084

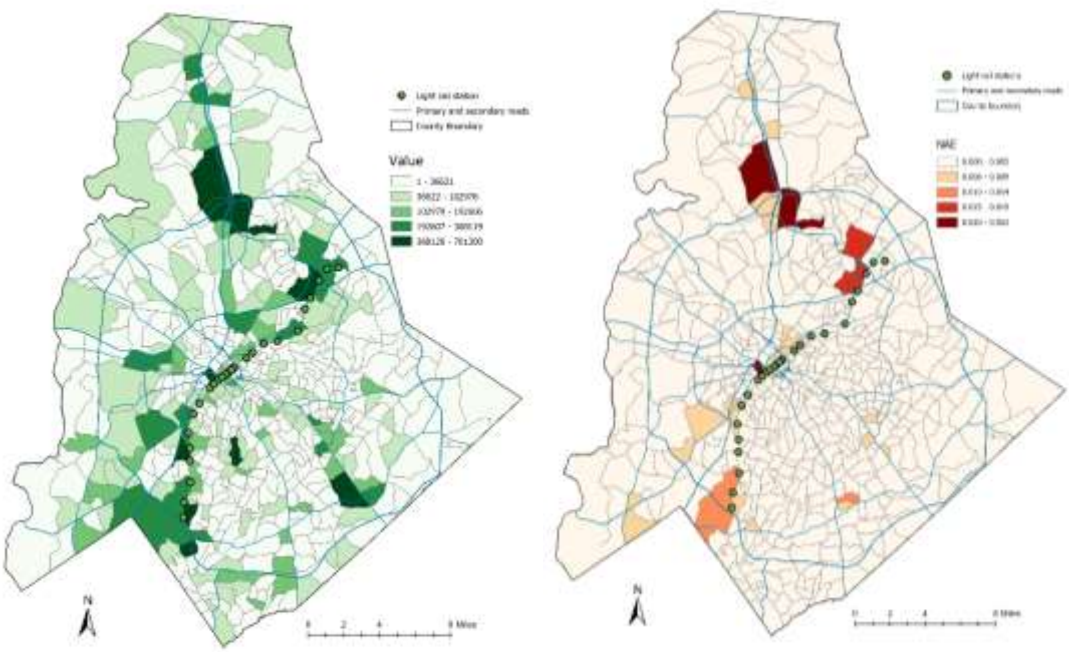
Having established the forecasting method performs very well overall for the study area, we can proceed with a closer examination of local patterns of forecasts across Mecklenburg County. In figure 4, we further illustrate the observed values (on the left) and predictive errors (on the right) using normalized absolute error (NAE) (normalized by the range of observed values) for the three outcomes variables of (a) employment, (b) business sales volume, and (c) productivity given by sales per employee in 2016. The absolute errors normalized by the mean of observed values are also mapped. They have a spatial distribution very similar to that of the absolute errors normalized by range. As a result, only the maps of the latter are reported here. As shown in figure 4, the majority of block groups have very low predictive errors. Only a few block groups have higher errors for all outcome variables. The maps of observed values show that these block groups with higher normalized errors all have larger observed values. It is natural to have larger errors in predicting larger values. Since these errors are normalized by a constant value (average observed values), the normalized errors naturally become larger. Even these block groups with higher errors actually have normalized errors under 0.33, 0.02, and 0.22 for employment, sales volume, and labor productivity, which shows performance above 67%. Furthermore, spatial dependence is tested on errors using the Global Moran's *I* statistic. Results show that there is no spatial autocorrelation with index values of 0.035, 0.064, and 0.0025 for employment, sales volume and productivity variables, respectively. This shows that there is no specific spatial pattern in the errors. There is no evidence of a spatial lag in the specification that would be omitted, nor are we omitting any predictor that would show a strong spatial tendency.

Observed values by block group Predictive errors by block group

(a)



(b)



(c)

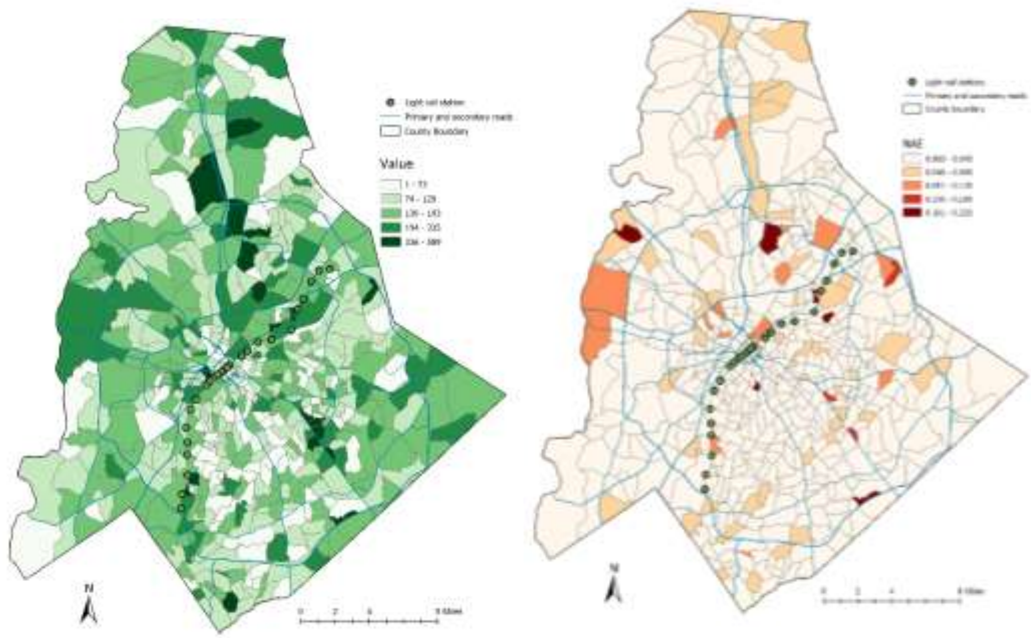


Figure 4. Observed values and model predictive errors (absolute error normalized by range) for (a) employment, (b) businesses sales volume, and (c) labor productivity in 2016

Figure 5 shows the statistical distribution of normalized predictive errors (normalized by the range of observed values) for each of the outcome variables for the year 2016. Comparing the histograms of three outcome variables shows that, for labor productivity, the errors are more skewed and there are more block groups in the larger error classes. This variable is calculated based on the other two variables (employment and sales). Therefore, this increases the level of uncertainty and noise in productivity. Therefore, predictions become somewhat more challenging.

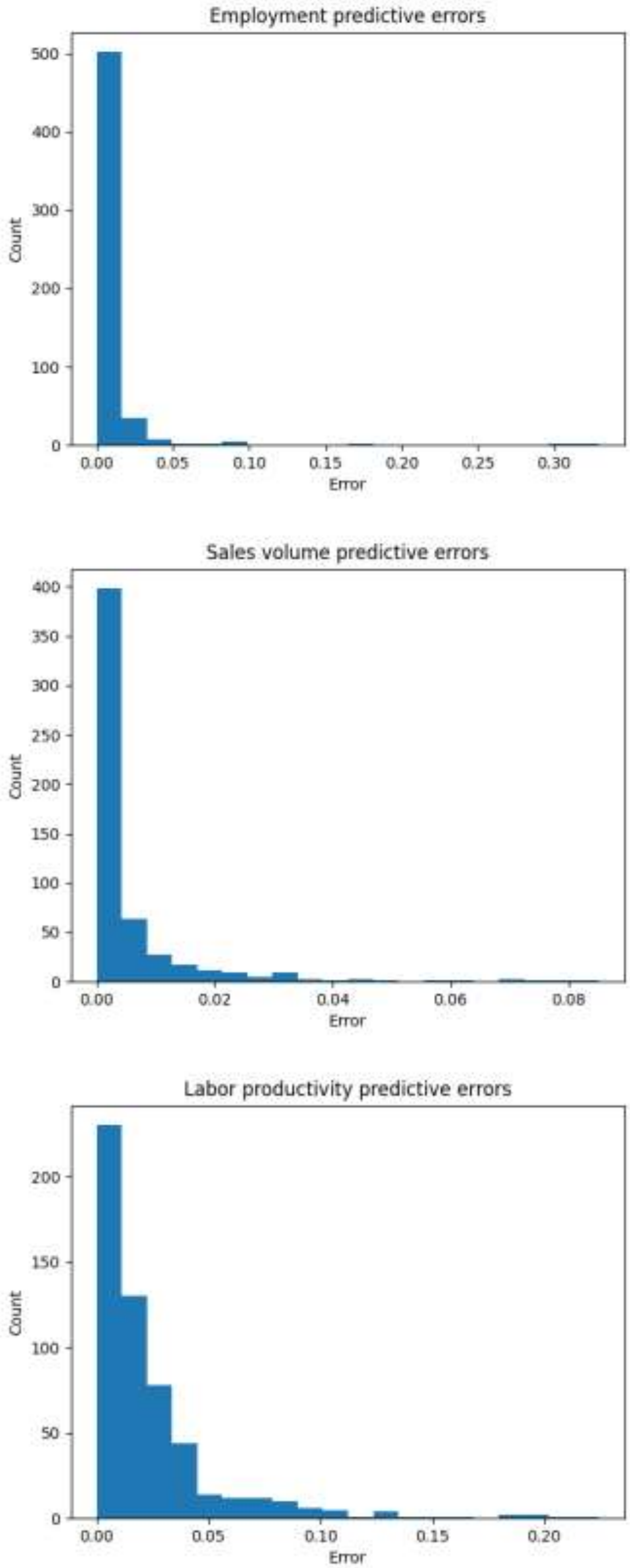


Figure 5. Statistical distribution of normalized predictive errors (normalized by the range of observed values) for the year 2016.

3.4. Discussion and conclusions

Forecasting economic development is critical as it informs both policy makers and business owners about the future prospects of different areas. By knowing with reasonable confidence about the future in advance, policy makers will be able to prepare if they need to investigate influential factors to set suitable policies for managing declining trends or encouraging growing trends in specific neighborhoods. Effective local forecasts may also be leveraged by public agencies to anticipate the needs for physical infrastructure (such as transportation) and human infrastructure (such as schools, day care centers or health facilities) in or near neighborhoods with fast evolving trends. Business owners and real estate developers can use forecasting to learn about locations with higher potential economic returns. Therefore, adopting more accurate forecasting models will result in higher efficiency in the urban economy. Performing the forecasts at fine geographic resolutions, such as neighborhoods, assists policy makers, community development managers, business owners, or real estate developers in making more accurate decisions.

Service industries consist of a considerable share of the economy, and their dynamics vary across different types, mainly whether they are in the group of business services or non-business services. As a result, in this study, three industries were selected and studied as non-business services together. These industries include retail, accommodation and food services, and other services with 2-digit NAICS codes of 44-45, 72, and 81, respectively.

In this study, three economic indicators of employment, business sales volume, and labor productivity were selected as outcome variables. In order to forecast these economic outcomes, covariates interrelated with economic dynamics were used in a multivariate model, including the sociodemographic, economic, built environment, and real estate investments. Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNN) was employed to handle the nonlinearities in the relationships. A spatiotemporal LSTM-RNN model was developed to forecast the outcomes for future years. In order to evaluate the model through out-of-sample testing, the dataset covering all block groups in Mecklenburg county, NC, for the years 2002-2018 was split into training and testing sets over temporal dimensions with proportions of 80% and 20%. The last three years are used for testing the model performance. Our model learns the relationships based on all census block groups together, and forecasts are done for each block group individually. Spatial autocorrelation is an important issue that must be taken into account when studying spatial data. Otherwise, in case of the presence of spatial dependence and failure to control for it, results might be biased. In this study, the spatial dependence is evaluated for the three outcome variables using Global Moran's I method (Moran, 1950) and no autocorrelation was found.

Our results indicate good performance of the model for the three outcome variables of employment, business sales volume, and labor productivity, with MAE values of 0.235, 0.206, and 0.164 as normalized by the mean of observed values; and MAE values of 0.011, 0.012, and 0.034 as normalized by the range of observed values. Normalization is done to have a better understanding of the error values given the wide range of observed values across all block groups. Looking into the results in figure 3, only a small number of block groups fall in the larger error classes. The majority of those are block groups with large observed values and it is natural to have larger predictive errors for larger observed values. One potential reason for the high predictive error of employment and sales in the block groups on the north corridor is the opening of the I-485 highway (intersecting with the I-

77 highway) in 2013. This significant change in the spatial structure of the city can lead to sudden jumps from historical trends. Additionally, as shown in the map and histogram of predictive errors for productivity (Figures 3 and 4), there are relatively more block groups in the top error class (largest errors). This is, to a large extent, due to higher uncertainty in this metric. The productivity metric is calculated based on two inputs of employment and sales volume, each of which brings some level of uncertainty and noise. As a result, the productivity metric becomes noisier and more challenging to predict.

As shown by maps of all three indicators, high economic outcomes are located along the 5 mobility axes aligned with the LYNX Blue Line, LYNX Blue Line Extension, proposed LYNX Red Line, proposed LYNX Silver Line, and the City LYNX Gold Line. These transit corridors were planned to integrate transportation and land use by enhancing accessibility to a mixture of land uses, including single-family and multifamily residential, retail and office (Charlotte-Area-Transit-System, 2019; City of Charlotte, 1998). One of the main objectives of transit-oriented developments along these corridors has been to promote economic development by bringing the origin-destinations of home, work, and retail closer together. As shown in maps of observed values (Figure 3), Mecklenburg County has large values of employment and sales volume along the transit corridors. In predictive modeling, it is expected to have larger errors for larger values. The maps of errors (Figure 3) indicates that the model developed in this article performs well as there are not many block groups with large errors along the transit corridors.

The ability to predict the future of economic outcomes in the county, particularly in competitively growing areas such as the mentioned transit corridors, would be tremendously informative to investors and business owners to locate areas with higher returns to investments. In addition, the fine geographic resolution of census block groups helps them to find their desired location in competitive areas with higher accuracy. The proposed spatiotemporal modeling framework has the capability of predicting outcomes in other industries as well.

One of our considerations for future work is to expand the study area to the metropolitan area to capture more interactions and dynamics in a functional boundary. In this study, as a result of data availability limitations, we had to confine the study to Mecklenburg County. Considering the majority of interactions and dynamics happening within Mecklenburg County, this study will not be problematic; however, incorporating interactions with neighboring counties may give a better and more comprehensive representation of reality. Surrounding counties do not make spatial data such as built environment-related shapefiles or building permit data available as Mecklenburg County does. This study calls for attention to the importance of data integration and building appropriate infrastructures to supply data for a functional region such as metropolitan areas, considering the matter of consistency.

In conclusion, although the causal relationships between different social, environmental, and economic factors are tremendously important, the literature must pay more attention to the future forecasts of economic outcomes in addition to the causal inference. The ability to predict the prospective outcomes benefits a variety of stakeholders, such as the public sector, urban planners and designers, business owners, and real estate investors, as well as the local residents. In this study, we developed a spatiotemporal deep learning model that predicts the economic outcomes, including employment, business sales volume, and labor productivity for three years at the fine geographic resolution of block groups. Having longer historical data is expected to enhance the model's predictive performance. Depending on data availability, we will consider using longer historical data of economic outcomes for future work as well as broader study area that naturally fits the extent of the metropolitan area. Furthermore, using information on the businesses' costs as an additional input would help us develop an economic indicator well suited to the profitability of businesses. As the most useful information from the business perspective, future profits can be studied in future work.

CHAPTER 4: FORECASTING THE KNOWLEDGE-INTENSIVE BUSINESS SERVICES PERFORMANCE IN LOCAL AREAS WITH A DEEP LEARNING FRAMEWORK

Abstract

As a result of the movement from an industrial to a knowledge-based economy in the last two decades, knowledge-intensive business services have been increasingly important in today's economy. These services make major contributions to the economy as a result of their impact on innovation and knowledge transfer. Therefore, it is critical to have a good understanding of their past and potential future trends. The literature on knowledge-intensive business services mainly investigates their causal impacts but there is a glaring lack of attention to their forecasting. In this study, we develop a multivariate spatiotemporal recurrent neural network long short-term memory model to forecast the three-year-ahead economic productivity of knowledge-intensive business services in Mecklenburg county, NC, at the fine resolution of census block groups. A considerable number of covariates are used to capture the relationships between different dimensions, such as sociodemographic, economic, and business characteristics, real estate investments, spatial structure, and the built environment. The model's performance is assessed using the median absolute error metric, which indicates small errors as 0.13 and 0.033, normalized by the mean and range of observed values, respectively. On this basis, the proposed framework for forecasting the business performance of small areas is found to be highly effective. With a deep learning model at its core, it is highly capable of capturing the complexities in high-dimensional data where traditional econometrics models cannot

perform well. Its distinctive features make it an advantageous analytical tool for a wide range of stakeholders.

Keywords

Knowledge-intensive business services, time series forecasting, deep learning, recurrent neural networks, long short-term memory, high-dimensional data, small area forecasting

4.1. Introduction

Given the economies of scale and agglomeration economies, cities have increasingly evolved into the economic engines of regions and nations. The agglomeration of economic entities in close proximities provides them with benefits such as access to a shared pool of labor and consumers, infrastructure, input materials and resources, and knowledge spillover, with lower transportation costs. These benefits encourage industries to cluster in the geographic space with each other or with other compatible industries (Rosenthal & Strange, 2004). Agglomeration forces dynamically impact changes in the economic growth of different industries in different areas of a city, region, or nation. As a result, studying dynamics in urban economies and looking into their futures is crucial for the economic and social vibrancy of neighborhood communities and of the cities they are part of, as well as to manage public resources needed to support sustained vibrancy and to avoid local economic instabilities.

Urban economic dynamics and agglomeration forces vary significantly across industries. One of these contrasted differences can be found in the service industries, where services can be rendered to customers who are either other businesses or non-business entities. Business services are less reliant on consumer foot traffic because of the differences on the demand side. Agglomeration is a function of accessibility and transportation costs (Chatman & Noland, 2011), however; business services are expected to be less sensitive to the level of accessibility of consumers in their location as a result of their lower dependence on foot traffic. In addition to their demand side and the role of accessibility, the contribution of business services and non-business services to the economy differs. Business services comprise a smaller share of the economy in terms of employment, with 43.42% of total employment in service industries in the US, while they cover a larger share in terms of wages, with 61.53% of total wages in service industries in the US in 2017, in comparison with the non-business services (US Census Bureau, 2017). Consistently with what shares of employment and wages indicate, business services jobs have higher wages in comparison with non-business services. They are mainly knowledge-intensive, meaning that their labor is highly skilled, and their activities include the generation, accumulation, and dissemination of knowledge (Muller & Doloreux, 2009). Therefore, it is necessary to study business services and non-business services separately.

Business services, particularly knowledge-intensive business services (KIBS), play a critical role in innovation processes through the generation and diffusion of knowledge (Muller & Doloreux, 2009; Toivonen, 2004). Industries in the KIBS can be considered bridges for innovation. Knowledge and innovation can be generated and disseminated in both: 1. industries within business services and 2. between the KIBS and non-service industries such as manufacturing. Part of this bridging function comes from the interactive and customer-related features of services provided by the KIBS for other KIBS or non-service industries (Muller & Doloreux, 2009). As a result, KIBS industries play a significant part in innovation and knowledge dissemination across a range of economic sectors.

Additionally, innovation is essential to the economy. Hausman & Johnston (2014) provide an explanation of how innovation contributes significantly to the economy. They contend that innovation is positively associated with job creation and employment. As a result of growth in employment and income, the spending of the population will increase, and this helps the economy to grow further.

Furthermore, discontinuous innovation is positively associated with economic stability. Economic downturns have provided evidence that, in comparison with other businesses, businesses with innovative strategies such as KIBS, suffer from less loss. Not only do the businesses with innovative strategies lose less in comparison with other businesses, but also, they can use this economic downturn as an opportunity to grow more. History has shown that many times, economic pressures motivate the assimilation of novel strategies and technologies. In addition, as unemployment increases, businesses have to compete less with others for employing highly skilled labor. They have the opportunity to hire a workforce with lower costs. Also, the unemployed population facing difficulty finding a new job may make efforts for entrepreneurship leading to novelty and innovation.

Muller & Doloreux (2009) mentioned the growth in KIBS activities as one of the most pivotal trends in economic evolutions. They argued for the increasing significance of the KIBS in the economy, particularly the economy of industrialized countries. As a result of advancements in technology and the dependence of business services on technology and their knowledge-intensive nature, these industries have been experiencing unprecedented competitive growth worldwide (Wirtz et al., 2015). Toivonen (2004) also pointed to the upward trends in the growth of the KIBS in the past and projected continuous expansion in the future, given the advancements in information technology.

In conclusion, as a result of the vital role of KIBS in innovation and the crucial function of innovation in the economy, as well as the upward trends in KIBS and innovations in recent decades, KIBS industries are important pillar of urban economies and forecasting their performance helps support the economic futures at the level of small areas like neighborhoods.

Despite the considerable attention paid to the business services dynamics in the literature on urban economies (Freel, 2016; Hertog, 2000; Miles, 2005; Muller & Doloreux, 2009), there is a gap in their approach to the study. The majority of the literature performs explanatory and confirmatory analyses on business services (Amara et al., 2009; Desmarchelier, 2013; Leiponen, 2006; Pina & Tether, 2016). Less attention has been paid to the forecasts of their outcomes and trends in the future. Given the ever-accelerating pace of advancements in technology and changes in dynamics in the economy, business services, particularly KIBS, are anticipated to expand more and make bigger contributions to the economy in the future (Toivonen, 2004). In addition, since businesses are the clients of business services and compose the majority of their demand side, input-output linkages constitute one of the key agglomeration drivers in business services. As a result, trends and dynamics in the business service sectors are highly interrelated. This interdependency and sensitivity to trends in other businesses lead to some level of uncertainty in economic outcomes of the business services.

Therefore, it is crucial to have a solid grasp of future patterns and potential growth or decline, given the essential role of business services in the economy in general, and uncertainty in economic outcomes across different industries. In order to understand these potential trends and reduce uncertainty, forecasting models can be highly beneficial. At various geographic levels of the country, the region, and the neighborhood, policy and decision making in economic problems might benefit from insights from future projections. The stated benefits of prediction models to policy making and decision making hold true for various areas such as land-use transportation interactions, economic development, real estate, affordable housing, etc. Furthermore, different scenarios and alternatives can be implemented in prediction models for evaluating the impact of socioeconomic interventions in policy-making and decision-making.

Recent advances in data availability and the growth of data-oriented approaches in research have paved the way for implementing forecasts for economic problems. High-resolution data in both spatial and temporal dimensions by fine industry disaggregation has been made available recently from resources such as the Longitudinal Employment Household Dynamics (LEHD) (U.S. Census Bureau, 2021c) and the US Businesses dataset from the data source Data Axle (Reference Solutions) (Data Axle, 2021) as well as other similar data providers. High-resolution economic data, along with sociodemographic data, can better represent socioeconomic phenomena. Meanwhile, recent advances in artificial intelligence (AI), such as machine learning (ML) and deep learning (DL), have enhanced our ability to capture the complex relationships between different dimensions of the complex and multifaceted area of urban dynamics and urban economics. Also, ML and DL approaches are not hampered by the assumptions in traditional statistics methods, such as linearity, normality, and multicollinearity. As a result, these methods are very well suited for investigating socioeconomic phenomena in a complex context, including non-linear relationships between a large number of covariates. The very recent trends in using ML and DL in economic forecasting (Arribas-Bel et al., 2021; Delmelle and Nilsson, 2021; Hatami et al., 2023; Rajesh et al., 2021; Ron-Ferguson et al., 2021) instead of the traditional econometric models are supporting evidence.

This study forecasts economic growth in business service industries in Mecklenburg County, NC, using historical data from 2002 to 2018. Using DL, the model makes three-year-ahead forecasts. The economic growth is measured by the productivity of KIBS at the neighborhood level. Productivity is measured by the annual ratio of the sales volume of businesses to their count of employees. Due to the expected relationships between the components of the urban economy, such as sociodemographic, economic, and business characteristics, real estate investments, spatial structure, and the built environment, multivariate modeling is utilized for forecasting.

One of the gaps in the literature on urban economies and economic growth, particularly economic growth of business services, pertains to the scale and unit of analysis. The majority of the literature studied the economy at the national or regional level (Billings & Johnson, 2016). However, as building blocks of regions, neighborhoods are important to be used for studying changes in the economy. Billings & Johnson (2016) explain the location decision of forms as a two-step process in which they first choose the state or region, and secondly, they choose the site or neighborhood within a city. There is heterogeneity in economic trends over different neighborhoods of a region. Given their socioeconomic characteristics, each neighborhood contributes differently to the region. In other words, agglomeration forces that impact the location prosperity of businesses vary over different neighborhoods (Billings & Johnson, 2016). As a result, intervention policies and economic strategies should be developed at the neighborhood level as well, in addition to the regional level. This study aggregates all the data to the census block groups representing neighborhoods.

4.2. Literature review

A variety of models have been developed to explain urban and regional economies from the classical models of agricultural land use developed by Von Thunen in which the value of land use is a function of the costs and ravenous of production and transportation to the market in those uses (Sinclair, 1967); central place theory of Walter Christaller and August Losch in which urban developments are structured around central places from which goods and services are produced and distributed to the surrounding areas in an isotropic plane (White et al., 2015); theory of industrial location developed by Alfred Weber in which industries are located in places where the costs of the raw materials, labor, production and distribution costs are minimized (White et al., 2015); monocentric city and bid rent theory developed by William Alonso (1960) in which land uses are formed as a function of distance to CBD to newer model of spatial interactions developed by Fotheringham and O'Kelly (1989) in which spatial locations interact which each other through flows of people, goods, services and information as a function of attractiveness of location and the cost of the distance between the origin and destination. These theories aimed to explain the formation of cities, regions, and the patterns within and between them through economic logic; however, the issue with those models is their static characteristics. They assume the city is a static phenomenon in a cross-section of time in which a state of equilibrium is assumed. On the contrary, cities and regions and their socio-economic characteristics are dynamic in nature and change over time (Batty, 2013).

In later years, with advancements in computational power and information technology, simulation models such as Cellular Automata (CA) were developed to explain how cities or regions evolve over time (White et al., 2015). Initially invented by Stanislaw Ulman and John von Neumann, CA models were aimed to simulate dynamics in complex systems. In these models, the area under study is divided into regular grid cells, and each cell is assigned and has the potential to be assigned a finite number of states, such as urban and rural, at different time steps. Then, over time cell states change based on some transition rules and the current states of cells and their neighboring cells (Mitchell, 2005). CA has had an evolutionary role in urban dynamics modeling; however, this model and its variants have had limitations, such as dependency on input parameters and the way the finite number of transition rules are defined that are supposed to represent complex reality. Also, transition rules are stationary over time which may not necessarily hold true for the reality of complex systems. The majority of studies using CA models have been dedicated to a limited number of applications. They have been mainly used for analyzing satellite imagery raster data and not sociodemographic vector data delineating irregular boundaries such as counties and census tracts. CA has been mostly used for modeling urban growth (Feng et al., 2020; Feng et al., 2011; Han et al., 2009; Liu et al., 2014; Moghadam & Helbich, 2013), urban land use change (Almeida et al., 2008; Almeida et al., 2003; Stevens & Dragićević, 2007; White & Engelen, 1993; White et al., 1997; Yang et al., 2008; Yang et al., 2012), ecology and land cover change (Fonstad, 2006; Gidey, 2017; Lu et al., 2019; Rimal et al., 2018). Usually, the model is used for discrete outcome variables and not continuous variables. For example, for predicting economic growth, CA models have been utilized for predicting different categories of land use, categories of developed versus undeveloped land (Liang, 2020), or different types of industries (Zoričak et al., 2021), and not economic growth in terms of socioeconomic variables such as income or businesses sales volume.

There is a considerable body of literature on economic growth in the US (e.g., Corrado et al., 2009; Fernald & Jones, 2014; Jones, 2002; Jorgenson & Fraumeni, 1992; Jorgenson & Stiroh, 2000; Jorgenson et al., 2016). Jorgenson and Stiroh (2000) pointed to the heterogeneity in economic growth trends and dynamics over different industries comprising the overall economy. They studied the economic growth in the US for years between 1958 to 1996 by decomposing the economy into 37 industries. They found that while the entire US economy experienced a growth of 0.45%, this growth varied over different industries, with some experiencing growth and some experiencing decline. Given the variation in economic trends over different industries, many research papers studied the economic growth of business services, their growth determinants, contributions to the overall economy, and the role of innovation in their growth, particularly KIBS (Brenner et al., 2018; Chadwick et al., 2008; Corrocher & Cusmano, 2014; Desmarchelier et al., 2013; Huggins, 2011; Miles, 2005).

Using firm-level survey data, (Corrocher, 2013) studied the growth patterns in knowledge-intensive business services in Lombardy, Italy. In this study, firm growth was measured by growth in sales as the outcome variable. Three factors of firm age, firm size, and innovation were investigated as drivers of economic growth. Corsi et al. (2019) divided the economy into two groups of KIBS and non-KIBS industries and investigated the contributions of each group to economic growth. They studied the contributions in samples of Italian and Spanish university spin-offs between 2005 to 2013. The study found that the positive impact of being KIBS on economic growth varied over study areas, with positive impacts only in Spain. Figueiredo & de Matos Ferreira (2020) Studied the prediction of the propensity to innovate in KIBS sectors in Brazil. In this study, KIBS were divided into professional KIBS and technological KIBS, and different sizes of firms, including small, medium, and large sizes were studied. Relationships between knowledge creation and transfer with innovation were investigated in KIBS.

However, the literature suffers from a lack of attention to future outcomes and predictions. Only a few studies explored future trend forecasts. Ahmadi et al. (2019) used the knowledge-based economy indicators for the years 1993 to 2013 to predict the economic growth in Iran for the years 2013 to 2020. They used machine learning models, including multilayer perceptron, adaptive neuro-fuzzy inference system, and gene expression programming for forecasting economic growth. Using time series data for years between 2000 and 2013 in Poland, Skórska (2015) investigated changes in KIBS economic growth and proposed a framework for forecasting future changes. Economic growth was measured by employment in these sectors.

4.3. Methodology

4.3.1. Workflow

In this paper, we use the basic foundation of economic models to generate a suitable variable for predicting the profitability of a business. Costs, revenues, and, therefore profits of a business (in economics, we use the term “firm”) can be represented by a production function in the context of our research. Assuming prices remain constant, a firm's profits are defined as a function of labor (number of employees) and capital (machinery, rent, etc.).

Also, by assuming that the firm is operating at its market equilibrium (normal state of operation), both labor and capital invested are functions of each other. Therefore, one variable can predict the other. In this state, the number of employees is an almost perfect instrument to measure every other cost since all other costs change with any change in the number of employees. Therefore, by using sales per employee, we can compare the profitability of businesses with each other since these businesses are working at the same price. The number of employees represents the total costs, and sales represent the total revenues. Sales per employee can therefore be used to describe profitability. In order to forecast KIBS productivity, a multivariate forecasting model is proposed. A number of variables are used as potential covariates labelled as economic factors, attributes of residents and workers, built environment characteristics, and building permits, the latter being a leading indicator of the economy. Furthermore, in order to resolve the high-dimensionality problem in the dataset, a dimensionality reduction technique is used before training the model. With dimensionality reduction, the computational time can be dramatically reduced without affecting the model performance in a meaningful way.

In this study, deep learning is used for forecasting the economic outcomes for knowledge-intensive business services. This method is used instead of traditional statistical methods due to its ability to capture complex and non-linear relationships. Traditional statistics methods have limitations in their assumptions, such as linearity, normality, and independence, that do not hold for complex phenomena such as dynamics in urban economies. Similarly, deep learning methods have advantages over simulation-based models, such as agent-based models that are common in the urban dynamics literature. These models are dependent on and very sensitive to input parameters that are specified by the modeler. On the contrary, these complex relationships between a large number of covariates will be learned from the data by deep learning methods. Deep learning models perform very well when dealing with high-dimensional and complex data due to the higher number of hidden layers in their architecture. Moreover, the recurrent neural network (RNN) adopted for this study is an advanced deep learning model for sequential data. In RNN models, each hidden layer neuron receives information from neurons in the previous layer and previous timesteps in that layer itself. In other words, it captures the relationships between variables considering their trends over time. The overall forecasting workflow proposed in this study is illustrated in figure 6.

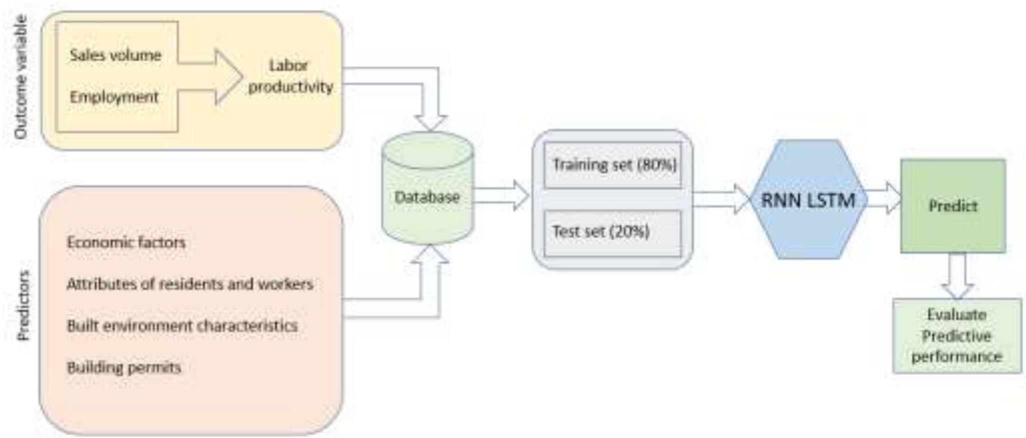


Figure 6. Proposed workflow for forecasting the business performance of KIBS

4.3.2. Knowledge-intensive business services

In this study, business productivity is predicted for specific business services. These business services are Professional, Scientific, and Technical Services, Management of Companies and Enterprises, and lastly, Administrative and Support and Waste Management and Remediation Services with 2-digit NAICS codes 54, 55, and 56, respectively. These industries are known as KIBS due to their role in the production and diffusion of knowledge in the economy (Muller & Zenker, 2001). KIBS have received significant attention in the urban economics literature as a result of their increasing leading role in overall economic growth and their direct impact on innovation (Muller & Zenker, 2001; Toivonen, 2004). The business productivity of the KIBS is predicted and tested using the annual historical data of years 2002 to 2018 at a fine geographic granularity within a mid-size city.

4.3.3. Study area

Mecklenburg County in North Carolina, USA, is the study area selected for this study. The county's economy for knowledge-intensive business services is studied for years 2002 to 2018 at the geographic level of census block groups representing neighborhoods (555 block groups).

Mecklenburg county is North Carolina's second most populous county (United States Census Bureau, 2021b) and has been experiencing upward trends in economic growth, with a 42% increase in employment between 2002 to 2018 (U.S. Census Bureau, 2021c). Economic growth has attracted populations from other counties and states, which has led to fast population growth over the recent decades. Its population grew by 58% between 2000 and 2019, with populations of 700,458 (U.S. Census Bureau; 2021a) and 1,110,356 (U.S. Census Bureau; 2021c), respectively.

The three industrial sectors comprising the KIBS industries have made significant contributions to Mecklenburg County's economy as they made 20.98% of total employment in 2018, with the Administrative and Support and Waste Management and

Remediation Services, and the Professional, technical, and scientific services in the top 5 industries (U.S. Census Bureau, 2021c).

The county experienced significant growth in the KIBS industries from 2002 to 2018, with a 90% increase in the Professional, technical, and scientific services (54 NAICS code), a 44% increase in Management of Companies and Enterprises (55 NAICS code), and 57% increase in Administrative and Support and Waste Management and Remediation Services (56 NAICS code) in the KIBS group, in terms of employment (U.S. Census Bureau, 2021c).

As a result of population and economic growth, urban plans were developed, such as the 2025 Transit/Land-Use plan aimed to integrate land use and transportation. This plan includes five transit corridors, two of which are the LYNX Blue line and the Blue Line Extension connecting the south of Charlotte to Uptown (city center) and to the University area in the northeast. These two lines started operation in 2007 and 2018, respectively. The implementation of the mentioned transit options and compact mixed-use developments in Charlotte led to changes in urban structure (from monocentric to polycentric) and the emergence of new employment centers in different areas of the county.

4.3.4. Data sources

Annual historical data of the KIBS sales volume are obtained from the US Businesses dataset from the data source Data Axle (Reference Solutions) (Data Axle, 2021). Reference Solutions provides the annual estimated sales volume for each business. In addition to the sales volume, this data source provides information on counts of employees by business and by year. A number of other employment information such as the employment by industry; age, income, education attainment, and race of workers; and business characteristics such as the firm's size and age by place of work and residence is collected from the Longitudinal Employer-Household Dynamics (U.S. Census Bureau, 2021c). The LEHD source provides employment data at the fine resolution of census block for years between 2002 and 2018. Data for real estate investments, known as a leading indicator of the economy (Strauss, 2013), are collected from the Basic Report for Posse Permits from the Mecklenburg County Open Data (2021). The POSSE dataset covers all the issued permits in Mecklenburg county. Data are provided by USDC code which shows the permit types, including new residential, new non-residential, commercial, demolitions, and additions. Moreover, built environment variables, such as proximity to Charlotte's Central Business District and rail transit proximity, are calculated using shapefiles collected from the Mecklenburg County's GIS Center (2021).

4.3.5. Description of data and methods

The study focuses on specific businesses with very similar characteristics. Table 11 shows the descriptive statistics of the outcome variable for the entire data history and by forecasting year. These businesses face a complex network of variables while operating in a major city such as Charlotte. No economic or logistic model can claim to be able to fully grasp the complexity of doing business in a major city. However, the multivariate model used in this paper can use all available socioeconomic variables. Variables indicating the agglomeration forces are included in the model. These variables are the number of employed residents showing the labor pooling, the number of businesses in the same industry, indicating colocation, and the location quotient showing localization. In addition, the educational attainment of workers is used as a proxy for the knowledge spillover effect as one of the agglomeration forces. Furthermore, neighborhood business characteristics such as their average age and size are included in the model covariates.

Table 11. Descriptive statistics of outcome variables

	Sales per employee ($1,000)			
	Average annual values for all years (2002-2018)	2016	2017	2018
count	9,435	555	555	555
mean	135.10	167.47	162.87	134.84
std	67.09	88.21	87.58	75.81
min	0	0	0	0
25%	97.60	109.29	109.34	83.37
50%	131.60	146.03	136.04	115.63
75%	164.00	215.41	209.88	172.62
max	630.62	637.45	634.55	482.96

For the purpose of forecasting, a number of business variables are used as potential predictors. Since business characteristics such as firms' age and size have an impact on their performance (Dunne & Hughes; 1994; Evans, 1987), the average age and size of firms in neighborhoods are included in the model. Furthermore, businesses' costs, such as wages, infrastructure, and facilities, have an impact on their growth or decline due to the economies of scale. This means that as the costs of a business or a neighborhood go up, the revenues will not necessarily go up linearly (white et al., 2015). As a result, employee wage as a business cost is taken into account in the predictive model to capture these non-linear relationships and scale effects. In addition, the input-output linkages between industries are one of the significant agglomeration forces in the business services economy (Billings & Johnson, 2016; Meliciani & Savona, 2015; Muller & Zenker, 2001). Businesses that utilize each other's inputs or outputs gain benefits from closer proximity as it reduces their transaction costs (Fujita et al., 1999). Leveraging economic data on other industries as producers and consumers as potential predictors, DL methods can capture complex relationships between industries. Furthermore, sociodemographic characteristics of neighborhoods, such as income, age, race, and educational attainment, are expected to have an impact on their economic outcomes. As a result, they are incorporated into the model as potential predictors.

Economic geography studies have found relationships between urban economies and the characteristics of the built environment, such as CBD proximity (Padilla & Eastlick, 2009; Wassmer, 2001), proximity to public transit (Bollinger & Ihlanfeldt, 1997; Cao & Porter-Nelson, 2016; Hess & Almeida, 2007; Knowles & Ferbrache, 2016; Levinson, 2010), Transit-Oriented Development (TOD) (Crampton, 2003; Credit, 2018; Renne, 2005), land-use mix (Iannillo & Fasolino, 2021; Wolf-Powers, 2005), real estate investments (Emil, 1991; Smith, 2009; Quigley, 2002). Given the available literature on the relationships between the built environment, urban spatial structure, and urban economies, this study uses built environment characteristics as predictors of economic outcomes.

Moreover, real estate investments have been known as leading indicators of the economy (Dua et al., 1999; Moore, 1983; Stock et al., 2008; Strauss, 2013). According to the literature, they have shown good predictive capability for economic outcomes such as growth or declines with a temporal gap. Therefore, in this study, real estate investments of different types, such as new constructions and renovations of residential and commercial properties, are utilized to predict the neighborhood’s economic outcomes. The building permits dataset is used with a one-year lag as they are said to be a leading indicator of the economy (Strauss, 2013).

In this study, the RNN long short-term memory (LSTM), as one of the most advanced sequential models (see Lipton et al., 2015; Mandic and Chambers, 2001; Medsker and Jain, 1999), is employed to forecast economic outcomes. Data from the years 2002 to 2018 are used for training and testing the model. The first 14 years (2002-2015) are used for training, and the last 3 years (2016-2018) are used for testing the model performance with 80%-20% training-testing set ratios. The model training is performed on all the block groups together as a data augmentation method. After the training phase, the model is tested by forecasting the economic outcomes for each census block group individually. After running the model, we used principle component analysis (PCA) to reduce the number of dimensions as our dataset has more than 300 features. We also ran the RNN LST model using the components instead of the full set of features. The PCA enhanced model performance and reduced the processing time significantly. In order to evaluate our model performance, we used evaluation metrics, including the mean absolute error (MAE). We used the outcome variable's logarithm before model training and the exponential function to make the forecasts in order to avoid negative predictions, which is a typical issue when predicting time series data. The model is developed in Python using the Keras library (Chollet, 2015).

4.4. Results

After training the model, we turn to predict the business performance of each neighborhood. Predictive errors are calculated to assess the model's predictive performance. Table 12 shows the mean absolute error (MAE) values by forecasting year and also averaged over all three forecasting years. Furthermore, the error measures are normalized by the mean and range of observed values to have a better understanding of their relative magnitude. The MAE metric is very informative for evaluating predictions.

But since we have a wide range of values across all block groups having a new metric relative to the observed values will be much more informative. On the other hand, percentage error, which is one of the most commonly used error metrics relative to the observed value, is not fully informative either since the practical magnitude of percentage error itself depends on the magnitude of the observed values. As a result, the normalized MAE metrics are used to have a better understanding of the errors. Table 12 shows 0.13 and 0.03 normalized error values. Also, as expected, errors by year show that the model performs better for the first and second years than the third year as this is prediction over a shorter term.

Table 12. Model's forecasting error values by year and averaged over all years

Evaluation metric	Forecast year	Average sales per employee ($1000)
MAE	All years	21.39
NMAE (by mean)	All years	0.13
NMAE (by range)	All years	0.033
MAE	2016	11.21
MAE	2017	10.97
MAE	2018	41.99
NMAE (by mean)	2016	0.06
NMAE (by mean)	2017	0.06
NMAE (by mean)	2018	0.31
NMAE (by range)	2016	0.02
NMAE (by range)	2017	0.02
NMAE (by range)	2018	0.09

We further mapped the predictive error values (Figure 7, on the right) for business services sales per employee in 2016. Maps for other years exhibit similar patterns and are therefore not shown here, in the interest of brevity. This figure shows that our model is able to predict the actual observed outcomes with a high degree of precision. The majority of block groups (more than 90% of block groups) have a normalized error value of less than

0.041, followed by values between 0.041 and 0.092 (7% of block groups). Only 5 out of the 555 block groups have larger normalized error values, between 0.093 and 0.354. Figure 8 illustrates the statistical distribution of these errors. The errors show a long tail pattern with a few block groups with large errors. The model produces high-accuracy predictions in areas with high productivity, such as the block groups in the North, Northwest, and Southwest of the county. Areas with large observed values are expected to produce larger errors. The model developed here provides good predictions in these areas in which larger errors are expected. North of the county, including areas near Lake Norman such as Cornelius, have been experiencing economic growth and the model forecasts the trend very well as error values are very low in these areas. Furthermore, spatial autocorrelation is tested using the global Moran's *I* index (Moran, 1950). Results show that there is no spatial autocorrelation, with a statistically non-significant index value of 0.002. Therefore, there is no evidence that spatial dependence would be present in the model, which could have taken the form of a spatial lag effect or of an omitted predictor with a strong spatial tendency.

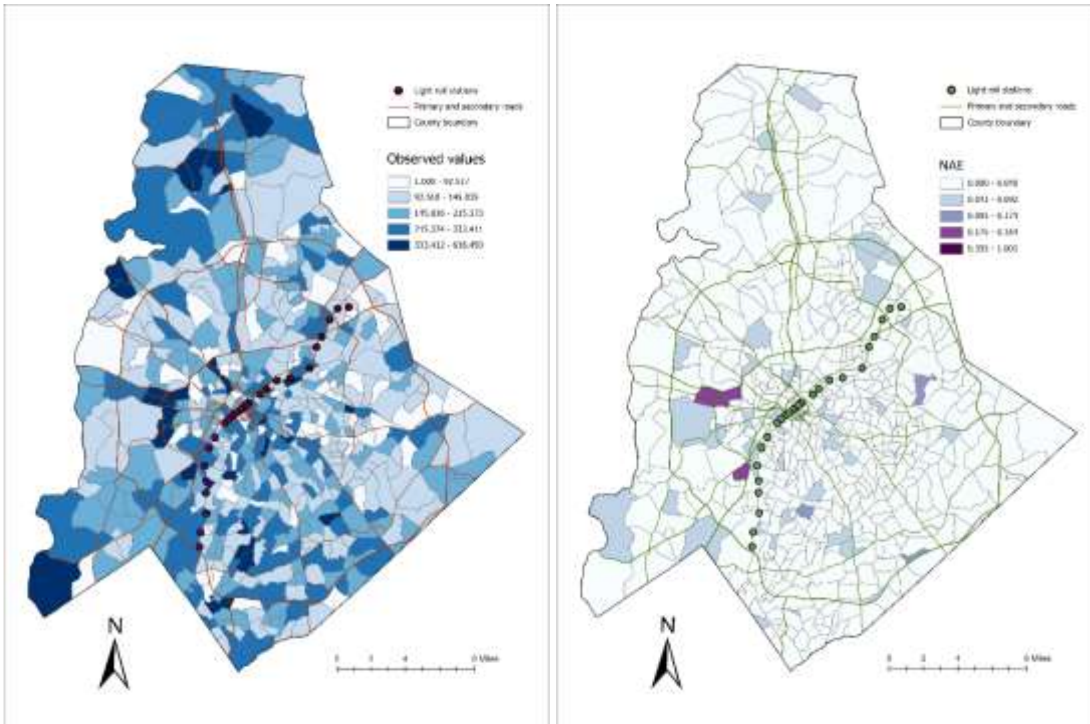


Figure 7. Map of observed businesses sales per employee in 2016 (on the left) and of absolute error values normalized by range of observed values (on the right)

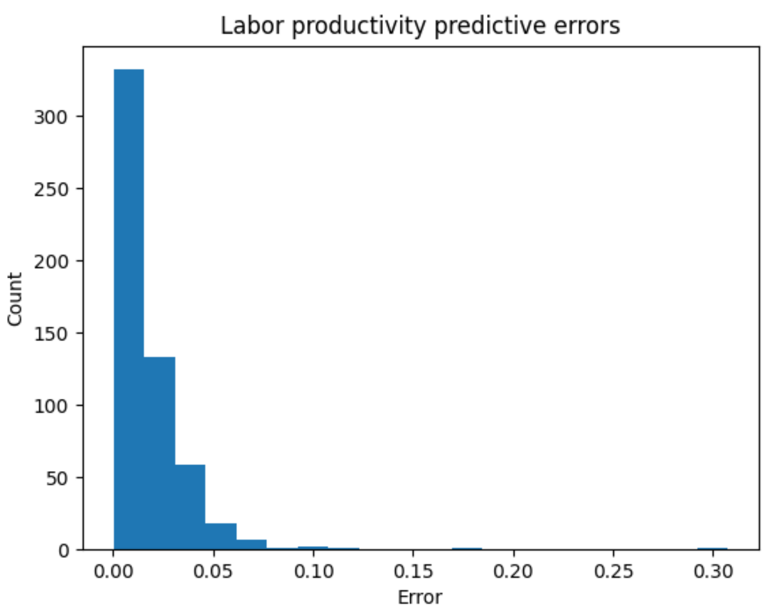


Figure 8. Histogram of normalized predictive errors (normalized by the range of observed values) for 2016

4.5. Discussion and conclusions

Any business, from small family-owned businesses to giant multinational corporations, is faced with a serious question when trying to open a new location. Where is the optimal location to start a business? Previous studies in economics and logistics have each attempted to incorporate a different set of variables to create a model that can find the optimal location to start a specific business within a limited context.

The common denominator of all these models is the most basic purpose of economics, which is to optimally allocate limited resources for maximum gains. In logistics, these limited resources are often space within a retail store (Campo et al., 2000). Consider a toy shop; there are a limited number of shelves available to the shop owner to stockpile toys to sell. Variables such as the distance to the nearest inventory, costs of transportation, number of customers, are then used to determine how many shelves need to be filled with what toys. We have the same basic setup in economics with a different class of variables. Instead of focusing on space, like in logistics, we focus on values (Agrawal & Smith, 2013). Resources are needed to be spent to produce a service that will yield a value equal to its price. Economic models deal with costs required to produce a service, which is deducted from the benefits gained from the sale of the said service. Consider the example of the toy shop mentioned above. A typical economic model considers the costs of labor, energy, rent, and the procurement of the toys themselves and then deducts these costs from the revenues generated by the sales. The remaining value will be the profits of the retail store. A retail store with the highest possible profit will be optimal. Therefore, by assuming that a business is operating with the objective of maximizing its net profits, we have decided to use efficiency as the key value in our study. We assumed that the firms operate at an equilibrium. Therefore, the optimal level of labor is utilized, which means that also the optimal level of capital is invested. At this stage, since both labor and capital move with each other, revenues per employee can be used as an indicator of efficiency. A firm that enjoys higher sales at the same labor level, therefore, enjoys a higher efficiency than other firms in the same industry. Given that the intra-urban scale of our study implies that labor costs (wages) and capital costs (energy costs, taxes, etc.) are the same for every firm in our sample, the only remaining variable that can affect this efficiency will be the location of the firm.

In this study, we proposed a new framework of multivariate spatiotemporal deep learning modeling using recurrent neural network long short-term memory aimed at forecasting the economic productivity of knowledge-intensive business services. We demonstrated this approach for small areas (census block groups) in Mecklenburg County, NC, for a three-year forecast period. The proposed model was trained and tested using a high-dimensional dataset that was created for this study. The mentioned dataset includes information on sociodemographic, economic, and business characteristics, real estate investments, spatial structure, and the built environment dimensions. The dataset was split into 80%-20% subsets for training and testing purposes, respectively. Moreover, PCA was employed as a dimensionality reduction technique given the large number of covariates that entered the analysis. The PCA technique reduced the processing time and enhanced the model's predictive performance significantly. The deep learning technique proposed in this paper enables us to circumvent the multitude of assumptions that other economic models need to be mathematically feasible. Our model uses a large number of available socioeconomic data and captures the complex relationships between covariates in this high-dimensional dataset.

The low predictive errors obtained in the case study demonstrate the power of our proposed model, even for small areas which are notoriously difficult to forecast with decent accuracy. Not only it predicts the growing trends in the key urban areas such as the CBD or transit corridors, but also it performs powerfully in predicting upward trends in less predictable areas such as the northern and southwestern edges of the county.

The nature of Charlotte itself and the vast investments in public transportation and light rail in the city make our findings even more important. Due to a combination of the overall tendency to reduce carbon emissions (Hoornweg et al., 2011), increase in population density (Sanchez et al., 2020), and shifts in public demand (Rappaport, 2008), we can expect that other cities across the US will try and copy Charlotte's strategy of expanding public transportation and investing in the required infrastructure (Ercan et al., 2016). This means that our model can be used to help businesses in those cities to have a better understanding of their future. It can also help policy makers more efficiently plan their city's expansion.

CHAPTER 5: CONCLUSIONS

Given the accelerating changes in today's world with all its positive and negative consequences and the role of urban areas in these changes, it is crucial to study urban dynamics. Moreover, it is essential to investigate urban phenomena considering their dynamic characteristics. The majority of the literature on urban studies takes cross-sectional approaches, neglecting the importance of the temporal dimension in urban patterns and processes. As a result, in order to address this gap, this study aims to take into account the dynamic nature by employing longitudinal research designs.

As a multifaceted topic, urban dynamics has been studied from a variety of aspects, such as environmental impacts, transportation/land-use integration, land-use and land cover change, and changes in the land price. This study explores two interrelated dimensions of urban dynamics: 1. built environment and travel behavior, and 2. economic growth.

Furthermore, as another gap in the literature on urban dynamics, the main focus has been on causal inference and explanatory analysis. Not until very recently have studies paid attention to predictions and temporal forecasts in urban dynamics. With regard to the fast pace changing dynamics, it is essential to gain insights into forthcoming patterns and processes. Therefore, this research investigates both the causal relationships and temporal forecasts in urban dynamics.

In the first chapter, the relationships between the built environment and commuting duration are studied. Based on the theoretical foundations of recent urban planning and design approaches such as new urbanism, smart growth, and transit-oriented development, It is hypothesized that built environments with higher degrees of density, diversity, design, destination accessibility, and lower distance to transit have shorter commuting duration. This study evaluates this hypothesis by taking two steps. First, the built environment is classified into four types of exurban, suburban, urban, and compact and transit-accessible developments (CTAD) employing Ward's clustering method. Second, the identified built environment types are used as independent variables to evaluate the differences in commuting duration between them. Using a number of spatial panel data models, we test the impact of the built environment on commuting duration, controlling for potential spatial dependences. The hypothesis is tested on Mecklenburg county block groups in the two years of 2000 and 2015. Results show that the built environment has a statistically significant impact on commuting duration. In both 2000 and 2015, CTAD areas had shorter commuting duration than the reference groups (exurban and suburban); however, comparing the two years together, commuting duration increased over time in these areas. Similarly, urban areas have shorter commuting duration than the reference groups. However, it is important to consider the practical significance in addition to the statistical significance. Results show that the magnitude of the impact is small. Hence, our findings are aligned with the ones that found weak or small impacts.

In the second chapter, a spatiotemporal multivariate forecasting model is developed using deep learning to forecast the three-year ahead economic performance of non-business services aggregated at the census block group geographic level. Three outcome variables are selected to measure the business performance; 1. employment, 2. business sales volume, and 3. labor productivity. The labor productivity variable is obtained by dividing the sales volume by the number of employees. A large number of covariates related to non-business service economic outcomes are incorporated into the model, such as accessibility by workers and consumers, distance to the CBD, and public transit, characteristics of businesses in the same or other industries, sociodemographic features of the residents and employees, and real estate investments.

The third chapter is focused on knowledge-intensive business services (KIBS), which have a significant role in today's economy. KIBS businesses make a substantial contribution to the economy due to their innovative role. In this chapter, a multivariate deep learning model is developed to forecast the three-year ahead productivity of these businesses at the census block group geographic level.

The developed models' predictive performances are evaluated using the mean absolute error (MAE) metric. All models performed well with low MAE values discussed in the results sections.

The deep learning employed in the second and third chapters helped overcome the high-dimensionality issue and handle the complexity of the data. We used the recurrent neural network long short-term memory (RNN LSTM) as an advanced deep learning method. RNN is capable of digesting complex relationships between a large number of variables due to its depth. Furthermore, it performs very well for sequential data as it uses values for previous time steps as an input in addition to covariate inputs.

In conclusion, both explanatory analysis and predictive modeling need more attention in urban dynamics. The multifaceted nature of urban dynamics with all its interrelated dimensions needs to be taken into account as well. Advanced ML and DL methods provide a promising bright future as a result of their capability to handle the discussed multidimensionality.

Addressing the discussed gaps in the literature, providing solutions to major issues, and developing high performance models, this study has been struggling with some limitations. One of the major limitations is the data availability. For the first article, accessing to data on attitudes and preferences of residents will help resolving the self-selection issue. Similarly, for articles two and three, having a longer period of historical data improves the training process. ML and DL models learn more complex relationships if seeing longer historical data. As a result, produce more accurate predicted values. This study considers continuing research in the future as more data becomes available. Another consideration for future work is including non-work travel behavior in article one as well. Noteworthy, the two types of travel (work and non-work) need to be studied separately as their dynamics are very different. Furthermore, Due to the complex relationships between different dimensions of the built environment-transportation relationships it is suggested to use methods such as structural equation modeling that handle multiple direction relationships between a number of variables.

References

1. Aburas, M. M., Ho, Y. M., Ramli, M. F., & Ash'aari, Z. H. (2016). The simulation and prediction of spatio-temporal urban growth trends using cellular automata models: A review. *International Journal of Applied Earth Observation and Geoinformation*, *52,* 380-389.

2. Ahmadi, M., Jafarzadeh-Ghoushchi, S., Taghizadeh, R., & Sharifi, A. (2019). Presentation of a new hybrid approach for forecasting economic growth using artificial intelligence approaches. *Neural Computing and Applications*, *31*(12), 8661-8680.

3. Allen, J., & Lu, K. (2003). Modeling and prediction of future urban growth in the Charleston region of South Carolina: a GIS-based integrated approach. *Conservation Ecology* 8(2): 2. URL: http://www.consecol.org/vol8/iss2/art2/

4. Almeida, C. D., Gleriani, J. M., Castejon, E. F., & Soares-Filho, B. S. (2008). Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science*, *22*(9), 943-963.

5. Almeida, C. M., Batty, M., Monteiro, A. M. V., Câmara, G., Soares-Filho, B. S., Cerqueira, G. C., & Pennachin, C. L. (2003). Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. *Computers, Environment and Urban Systems*, *27*(5), 481-509.

6. Alonso, W. (1960, January). A theory of the urban land market. Bobbs-Merrill Company, College Division.

7. Amara, N., Landry, R., & Doloreux, D. (2009). Patterns of innovation in knowledge-intensive business services. *The Service Industries Journal*, *29*(4), 407-430.

8. American Community Survey (2019). Retrieved from https://www.census.gov/topics/employment/commuting/guidance/commuting/acs-1yr.html

9. Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis*, *20*(1), 1-17.

10. Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis*, *27*(2), 93-115.

11. Anselin, L., & Bera, A. K. (1998). Introduction to spatial econometrics. *Handbook of applied economic statistics*, *237*.

12. Arribas-Bel, D., Garcia-López, M. À., & Viladecans-Marsal, E. (2021). Building (s and) cities: Delineating urban areas with a machine learning algorithm. *Journal of Urban Economics*, *125*, 103217.

13. Bagley, M. N., & Mokhtarian, P. L. (2002). The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *The Annals of regional science*, *36*(2), 279-297.

14. Barrington-Leigh, C., & Millard-Ball, A. (2015). A century of sprawl in the United States. *Proceedings of the National Academy of Sciences*, *112*(27), 8244-8249.

15. Batty, M. (2013). *The new science of cities*. MIT press.

16. Batty, M., & Torrens, P. M. (2001). Modelling complexity: the limits to prediction. *Cybergeo: European Journal of Geography*. Retrieved from https://journals.openedition.org/cybergeo/1035

17. Berg, N. (2014). Success from satisficing and imitation: Entrepreneurs' location choice and implications of heuristics for local economic development. *Journal of Business Research*, *67*(8), 1700-1709.

18. Billings, S. B., & Johnson, E. B. (2016). Agglomeration within an urban area. *Journal of Urban Economics*, *91*, 13-25.

19. Boarnet, M., & Crane, R. (2001). The influence of land use on travel behavior: specification and estimation strategies. *Transportation Research Part A: Policy and Practice*, *35*(9), 823-845.

20. Bollinger, C. R., & Ihlanfeldt, K. R. (1997). The impact of rapid rail transit on economic development: The case of Atlanta's MARTA. *Journal of Urban Economics*, *42*(2), 179-204.

21. Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, *22*(3), 547-581.

22. Brenner, T., Capasso, M., Duschl, M., Frenken, K., & Treibich, T. (2018). Causal relations between knowledge-intensive business services and regional employment growth. *Regional Studies*, *52*(2), 172-183.

23. Burkom, H. S., Murphy, S. P., & Shmueli, G. (2007). Automated time series forecasting for biosurveillance. *Statistics in Medicine*, *26*(22), 4202-4218.

24. Cao, X. J., & Porter-Nelson, D. (2016). Real estate development in anticipation of the Green Line light rail transit in St. Paul. *Transport Policy*, *51*, 24-32.

25. Cao, X. J., Mokhtarian, P. L., & Handy, S. L. (2009). The relationship between the built environment and nonwork travel: A case study of Northern California. *Transportation Research Part A: Policy and Practice*, *43*(5), 548-559.

26. Cao, X. J., Xu, Z., & Fan, Y. (2010). Exploring the connections among residential location, self-selection, and driving: Propensity score matching with multiple treatments. *Transportation research part A: policy and practice*, *44*(10), 797-805.

27. Cao, X., Mokhtarian, P. L., & Handy, S. L. (2007). Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach. *Transportation*, *34*(5), 535-556.

28. Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. *Transport reviews*, *29*(3), 359-395.

29. Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D: Transport and Environment*, *7*(4), 265-284.

30. Cervero, R., & Duncan, M. (2006). 'Which Reduces Vehicle Travel More: Jobs-Housing Balance or Retail-Housing Mixing?. *Journal of the American planning association*, *72*(4), 475-490.

31. Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation research. Part D, Transport and environment*, *2*(3), 199-219.

32. Cervero, R., & Murakami, J. (2010). Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. *Environment and planning A*, *42*(2), 400-418.

33. Chadwick, A., Glasson, J., & Smith, H. L. (2008). Employment growth in knowledge-intensive business services in Great Britain during the 1990s—variations at the regional and sub-regional level. *Local Economy*, *23*(1), 6-18.

34. Charlotte-Area-Transit-System. 2019. (2030) Transit System Corridor Plan. Available online: **https://charlottenc.gov/cats/transit-planning/2030-plan/Documents/2030_Transit_Corridor_System_Plan.pdf** (accessed on 10 October 2020).

35. Chatman, D. G., & Noland, R. B. (2011). Do public transport improvements increase agglomeration economies? A review of literature and an agenda for research. *Transport Reviews*, *31*(6), 725-742.

36. Chen, C., Gong, H., & Paaswell, R. (2008). Role of the built environment on mode choice decisions: additional evidence on the impact of density. *Transportation*, *35*(3), 285-299.

37. Chollet, F. (2015) keras, GitHub. https://github.com/fchollet/keras

38. Chong, S. K., Bahrami, M., Chen, H., Balcisoy, S., & Bozkaya, B. (2020). Economic outcomes predicted by diversity in cities. *EPJ Data Science*, *9*(1), 17.

39. City of Charlotte. (1998). 2025 Integrated Transit/Land-Use Plan. Retrieved from https://charlottenc.gov/cats/transit-planning/NorthWest%20RFP/2025%20Transit%20Land-Use%20Plan-Final.pdf

40. City of Charlotte. (2020). Charlotte Future 2040 Comprehensive Plan. Retrieved from https://www.cltfuture2040plan.com/sites/all/themes/custom/smoky_hollow/docs/CF2040_Public-Review-Draft-Plan_Web.pdf

41. Clarke, K. C., & Gaydos, L. J. (1998). Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, *12*(7), 699-714.

42. Cliff, A. D., & Ord, K. (1970). Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography*, *46*(sup1), 269-292.

43. Combes, P. P., Gobillon, L., & Zylberberg, Y. (2021). Urban economics in a historical perspective: Recovering data with machine learning. *Regional Science and Urban Economics*, 103711.

44. Corrado, C., Hulten, C., & Sichel, D. (2009). Intangible capital and US economic growth. *Review of income and wealth*, *55*(3), 661-685.

45. Corrocher, N., & Cusmano, L. (2014). The 'KIBS engine'of regional innovation systems: Empirical evidence from European regions. *Regional Studies*, *48*(7), 1212-1226.

46. Corrocher, N., Cusmano, L., & Lenzi, C. (2013). Growth in knowledge-intensive business services: Evidence from Lombardy. *Industry and Innovation*, *20*(6), 563-584.

47. Corsi, C., Prencipe, A., Rodríguez-Gulías, M. J., Rodeiro-Pazos, D., & Fernández-López, S. (2019). Growth of KIBS and non-KIBS firms: evidences from university spin-offs. *The Service Industries Journal*, *39*(1), 43-64.

48. Crampton, G. (2003). Economic development impacts of urban rail transport.

49. Crane, R. (2000). The influence of urban form on travel: an interpretive review. *Journal of Planning literature, 15*(1), 3-23.

50. Crane, R., & Crepeau, R. (1998). Does neighborhood design influence travel?: A behavioral analysis of travel diary and GIS data. *Transportation Research Part D: Transport and Environment*, *3*(4), 225-238.

51. Credit, K. (2018). Transit-oriented economic development: The impact of light rail on new business starts in the Phoenix, AZ Region, USA. *Urban Studies*, *55*(13), 2838-2862.

52. Data Axle. (2021). US Businesses. *Reference Solutions*. Retrieved from https://www.referenceusa.com

53. Day, K. (2003). New urbanism and the challenges of designing for diversity. *Journal of Planning Education and Research*, *23*(1), 83-95.

54. Delmelle, E. C. (2022). GIScience and neighborhood change: Toward an understanding of processes of change. *Transactions in GIS*, *26*(2), 567-584.

55. Delmelle, E. C., & Nilsson, I. (2021). The language of neighborhoods: A predictive-analytical framework based on property advertisement text and mortgage lending data. *Computers, Environment and Urban Systems*, *88*, 101658.

56. Delmelle, E. C., Zhou, Y., & Thill, J. C. (2014). Densification without growth management? Evidence from local land development and housing trends in Charlotte, North Carolina, USA. *Sustainability*, *6*(6), 3975-3990.

57. Desmarchelier, B., Djellal, F., & Gallouj, F. (2013). Knowledge intensive business services and long term growth. *Structural Change and Economic Dynamics*, *25*, 188-205.

58. Dolega, L., & Lord, A. (2020). Exploring the geography of retail success and decline: A case study of the Liverpool City Region. *Cities*, *96*, 102456.

59. Dua, P., Miller, S. M., & Smyth, D. J. (1999). Using leading indicators to forecast US home sales in a Bayesian vector autoregressive framework. *The Journal of Real Estate Finance and Economics*, *18*(2), 191-205.

60. Dunne, P., & Hughes, A. (1994). Age, size, growth and survival: UK companies in the 1980s. *The Journal of Industrial Economics*, 115-140.

61. Elhorst, J. P. (2014). Matlab software for spatial panels. *International Regional Science Review*, *37*(3), 389-405.

62. Emil, M. (1991). Forecasting demand for commercial real estate based on the economic fundamentals of US metro markets. *Journal of Real Estate Research*, *6*(3), 251-265.

63. Etminani-Ghasrodashti, R., & Ardeshiri, M. (2015). Modeling travel behavior by the structural relationships between lifestyle, built environment and non-working trips. *Transportation Research Part A: Policy and Practice*, *78*, 506-518.

64. Evans, D. S. (1987). The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *The journal of industrial economics*, 567-581.

65. Ewing, R., & Cervero, R. (2001). Travel and the built environment: a synthesis. *Transportation research record*, *1780*(1), 87-114.

66. Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American planning association*, *76*(3), 265-294.

67. Ewing, R., Tian, G., Goates, J. P., Zhang, M., Greenwald, M. J., Joyce, A., ... & Greene, W. (2015). Varying influences of the built environment on household travel in 15 diverse regions of the United States. *Urban Studies*, *52*(13), 2330-2348.

68. Feng, Y., & Tong, X. (2020). A new cellular automata framework of urban growth modeling by incorporating statistical and heuristic methods. *International Journal of Geographical Information Science*, *34*(1), 74-97.

69. Feng, Y., Liu, Y., Tong, X., Liu, M., & Deng, S. (2011). Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. *Landscape and Urban Planning*, *102*(3), 188-196.

70. Fernald, J. G., & Jones, C. I. (2014). The future of US economic growth. *American economic review*, *104*(5), 44-49.

71. Figueiredo, R., & de Matos Ferreira, J. J. (2020). Spinner model: prediction of propensity to innovate based on knowledge-intensive business services. *Journal of the Knowledge Economy*, *11*(4), 1316-1335.

72. Fonstad, M. A. (2006). Cellular automata as analysis and synthesis engines at the geomorphology–ecology interface. *Geomorphology*, *77*(3-4), 217-234.

73. Fotheringham, A. S., & O'Kelly, M. E. (1989). *Spatial interaction models: formulations and applications* (Vol. 1, p. 989). Dordrecht: Kluwer Academic Publishers.

74. Freel, M. (2016). *Knowledge-intensive business services: geography and innovation*. Routledge.

75. Fujita, M., Krugman, P. R., & Venables, A. (1999). *The spatial economy: Cities, regions, and international trade*. MIT press.

76. Ghose, A. (2019). Problems of endogeneity in social science research. In *Research Methodology for Social Sciences* (pp. 218-252). Routledge India.

77. Gidey, E., Dikinya, O., Sebego, R., Segosebe, E., & Zenebe, A. (2017). Cellular automata and Markov Chain (CA_Markov) model-based predictions of future land use and land cover scenarios (2015–2033) in Raya, northern Ethiopia. *Modeling Earth Systems and Environment*, *3*(4), 1245-1262.

78. Giuliano, G., Kang, S., & Yuan, Q. (2019). Agglomeration economies and evolving urban form. *The Annals of Regional Science*, *63*(3), 377-398.

79. Glaeser, E. L., Kim, H., & Luca, M. (2018, May). Nowcasting gentrification: using yelp data to quantify neighborhood change. In *AEA Papers and Proceedings* (Vol. 108, pp. 77-82).

80. Graves, A., & Graves, A. (2012). Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, 37-45.

81. Grekousis, G. (2019). Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis. *Computers, Environment and Urban Systems*, *74*, 244-256.

82. Guerra, E. (2014). The built environment and car use in Mexico City: Is the relationship changing over time?. *Journal of Planning Education and Research*, *34*(4), 394-408.

83. Han, J., Hayashi, Y., Cao, X., & Imura, H. (2009). Application of an integrated system dynamics and cellular automata model for urban growth assessment: A case study of Shanghai, China. *Landscape and urban planning*, *91*(3), 133-141.

84. Handy, S. (1993). Regional versus local accessibility: Implications for nonwork travel.

85. Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment*, *10*(6), 427-444.

86. Handy, S., Cao, X., & Mokhtarian, P. L. (2006). Self-selection in the relationship between the built environment and walking: Empirical evidence from Northern California. *Journal of the American Planning Association*, *72*(1), 55-74.

87. Hanlon, W., & Miscio, A. (2014). *Agglomeration: A dynamic approach* (No. w20728). National Bureau of Economic Research.

88. Hatami, F., Rahman, M. M., Nikparvar, B., & Thill, J. C. (2023). Non-linear associations between the urban built environment and commuting modal split: A random forest approach and SHAP evaluation. *IEEE Access*. vol. 11, pp. 12649-12662, 2023, doi: 10.1109/ACCESS.2023.3241627.

89. Hausman, A., & Johnston, W. J. (2014). The role of innovation in driving the economy: Lessons from the global financial crisis. *Journal of Business Research*, *67*(1), 2720-2726.

90. He, J., Li, X., Yao, Y., Hong, Y., & Jinbao, Z. (2018). Mining transition rules of cellular automata for simulating urban expansion by using the deep learning

techniques. *International Journal of Geographical Information Science*, *32*(10), 2076-2097.

91. Hertog, P. D. (2000). Knowledge-intensive business services as co-producers of innovation. *International journal of innovation management*, *4*(04), 491-528.

92. Hess, D. B., & Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. *Urban studies*, *44*(5-6), 1041-1068.

93. Hess, D. B., & Lombardi, P. A. (2004). Policy support for and barriers to transit-oriented development in the inner city: Literature review. *Transportation Research Record*, *1887*(1), 26-33.

94. Huggins, R. (2011). The growth of knowledge-intensive business services: Innovation, markets and networks. *European Planning Studies*, *19*(8), 1459-1480.

95. Iannillo, A., & Fasolino, I. (2021). Land-Use Mix and Urban Sustainability: Benefits and Indicators Analysis. *Sustainability*, *13*(23), 13460.

96. Izanloo, A., Rafsanjani, A. K., & Ebrahimi, S. P. (2017). Effect of commercial land use and accessibility factor on traffic flow in Bojnourd. *Journal of Urban Planning and Development*, *143*(2), 05016016.

97. Jacobs, J. (1961). *The Death and Life of Great American Cities*.

98. Jin, Y., Denman, S., Deng, D., Rong, X., Ma, M., Wan, L., ... & Long, Y. (2017). Environmental impacts of transformative land use and transport developments in the Greater Beijing Region: Insights from a new dynamic spatial equilibrium model. *Transportation Research Part D: Transport and Environment*, *52*, 548-561.

99. John, B. (1996). Mass transportation, apartment rent and property values. *Journal of Real Estate Research*, *12*(1), 1-8.

100. Jones, C. I. (2002). Sources of US economic growth in a world of ideas. *American economic review*, *92*(1), 220-239.

101. Jorgenson, D. W., & Fraumeni, B. M. (1992). Investment in education and US economic growth. *The Scandinavian Journal of Economics*, S51-S70.

102. Jorgenson, D. W., & Stiroh, K. J. (2000). US economic growth at the industry level. *American Economic Review*, *90*(2), 161-167.

103. Jorgenson, D., Gollop, F. M., & Fraumeni, B. (2016). *Productivity and US economic growth*. Elsevier.

104. Khattak, A. J., & Rodriguez, D. (2005). Travel behavior in neo-traditional neighborhood developments: A case study in USA. *Transportation Research Part A: Policy and Practice, 39*(6), 481-500.

105. Khusni, U., Dewangkoro, H. I., & Arymurthy, A. M. (2020, September). Urban area change detection with combining CNN and RNN from sentinel-2 multispectral remote sensing data. In *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)* (pp. 171-175). IEEE.

106. Kloosterman, R. C., & Musterd, S. (2001). The polycentric urban region: towards a research agenda. *Urban Studies, 38*(4), 623-633.

107. Knowles, R. D., & Ferbrache, F. (2016). Evaluation of wider economic impacts of light rail investment on cities. *Journal of Transport Geography, 54*, 430-439.

108. Koebel, C. T. (2014). *Analyzing neighborhood retail and service change in six cities*. Virginia Center for Housing Research, Virginia Tech, Virginia.

109. Kourtit, K., Pele, M. M. M., Nijkamp, P., & Pele, D. T. (2021). Safe cities in the new urban world: A comparative cluster dynamics analysis through machine learning. *Sustainable Cities and Society*, *66*, 102665.

110. Krueger, R., & Gibbs, D. (2008). 'Third wave'sustainability? Smart growth and regional development in the USA. *Regional Studies*, *42*(9), 1263-1274.

111. Leck, E. (2006). The impact of urban form on travel behavior: A meta-analysis. *Berkeley Planning Journal*, *19*(1).

112. Leiponen, A. (2006). Managing knowledge for innovation: the case of business-to-business services. *Journal of product innovation management*, *23*(3), 238-258.

113. Levinson, D. (2010). Economic development impacts of high-speed rail. *2010-06-05)[2013-04-01]. Nexus Research Group. Minneso-ta: University of Minnesota, 2010. http://nexus. urnn. edu/Papers/Eco-nomicDevelopmentlmpactsOfHSR. pdf*.

114. Liang, X., Liu, X., Chen, G., Leng, J., Wen, Y., & Chen, G. (2020). Coupling fuzzy clustering and cellular automata based on local maxima of development potential to model urban emergence and expansion in economic development zones. *International Journal of Geographical Information Science*, *34*(10), 1930-1952.

115. Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.

116. Liu, X., Ma, L., Li, X., Ai, B., Li, S., & He, Z. (2014). Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. *International Journal of Geographical Information Science*, *28*(1), 148-163.

117. Lu, Y., Wu, P., Ma, X., & Li, X. (2019). Detection and prediction of land use/land cover change using spatiotemporal data fusion and the Cellular Automata–Markov model. *Environmental monitoring and assessment*, *191*(2), 68.

118. Lv, Y., Lan, Z., Kan, C., & Zheng, X. (2021). Polycentric urban development and its determinants in China: A geospatial big data perspective. *Geographical Analysis*, *53*(3), 520-542.

119. Ma, S., Li, S., & Zhang, J. (2023). Spatial and deep learning analyses of urban recovery from the impacts of COVID-19. *Scientific Reports*, *13*(1), 2447.

120. Mandic, D., & Chambers, J. (2001). *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. Wiley.

121. Manson, S., Schroeder, J., Van Riper, D., & Ruggles, S. (2017). Ipums national historical geographic information system: Version 12.0 [database]. *Minneapolis: University of Minnesota*, *39*. Data retrieved from https://data2.nhgis.org/main

122. Manuele, F. A. (2009). Leading & lagging indicators. *Professional Safety*, *54*(12), 28.

123. Manville, M. (2017). Travel and the built environment: time for change. *Journal of the American Planning Association*, *83*(1), 29-32.

124. Martin, R., & Sunley, P. (1998). Slow convergence? The new endogenous growth theory and regional development. *Economic Geography*, *74*(3), 201-227.

125. Martínez, L. M., & Viegas, J. M. (2009). Effects of transportation accessibility on residential property values: Hedonic Price Model in the Lisbon, Portugal, metropolitan area. *Transportation Research Record*, *2115*(1), 127-137.

126. Mecklenburg County GIS Center. (2017). Data retrieved from https://www.mecknc.gov/LUESA/GIS/Pages/Default.aspx?src=hh

127. Mecklenburg County GIS Center. (2021). Data retrieved from https://www.mecknc.gov/LUESA/GIS/Pages/Default.aspx?src=hh

128. Mecklenburg County Open Data. (2021). *Basic Report for Posse Permits*. Retrieved from http://meck-ent-power-bi-reports-public web.azurewebsites.net/Viewer?groupId=2ca2c01c-5658-4ee2-8a4f-5bdda8be94b7&reportId=3b014920-672d-43cd-b920-201adde4be71

129. Medsker, L., & Jain, L. C. (Eds.). (1999). *Recurrent neural networks: design and applications*. CRC press.

130. Meliciani, V., & Savona, M. (2015). The determinants of regional specialisation in business services: agglomeration economies, vertical linkages and innovation. *Journal of Economic Geography*, *15*(2), 387-416.

131. Miles, I. (2005). Knowledge intensive business services: prospects and policies. *foresight*.

132. Mitchell, M. (2005). Computation in Cellular Automata: A Selected Review.

133. Moghadam, H. S., & Helbich, M. (2013). Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains-cellular automata urban growth model. *Applied Geography*, *40*, 140-149.

134. Mohammad, S. I., Graham, D. J., Melo, P. C., & Anderson, R. J. (2013). A meta-analysis of the impact of rail projects on land and property values. *Transportation Research Part A: Policy and Practice*, *50*, 158-170.

135. Mokhtarian, P. L., & Cao, X. (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological*, *42*(3), 204-228.

136. Moore, G. H. (1983). Why the leading indicators really do lead. In *Business Cycles, Inflation, and Forecasting, 2nd edition* (pp. 339-352). Ballinger.

137. Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17-23.

138. Muller, E., & Doloreux, D. (2009). What we should know about knowledge-intensive business services. *Technology in society*, *31*(1), 64-72.

139. Muller, E., & Zenker, A. (2001). Business services as actors of knowledge transformation: the role of KIBS in regional and national innovation systems. *Research policy*, *30*(9), 1501-1516.

140. Nasri, A., & Zhang, L. (2012). Impact of metropolitan-level built environment on travel behavior. *Transportation Research Record*, *2323*(1), 75-79.

141. Nasri, A., & Zhang, L. (2015). Assessing the impact of metropolitan-level, county-level, and local-level built environment on travel behavior: Evidence from 19 US urban areas. *Journal of Urban Planning and Development*, *141*(3), 04014031.

142. Nikparvar, B., Rahman, M. M., Hatami, F., & Thill, J. C. (2021). Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network. *Scientific Reports*, *11*(1), 21715.

143. Openshaw, S. (1979). A million or so correlation coefficients, three experiments on the modifiable areal unit problem. In N. Wrigley, ed. *Statistical Applications in the Spatial Sciences*, 127–144. London: Pion.

144. Padilla, C., & Eastlick, M. A. (2009). Exploring urban retailing and CBD revitalization strategies. *International Journal of Retail & Distribution Management*.

145. Pan, X., Liu, Z., He, C., & Huang, Q. (2022). Modeling urban expansion by integrating a convolutional neural network and a recurrent neural network. *International Journal of Applied Earth Observation and Geoinformation*, *112*, 102977.

146. Park, S., Jeon, S., Kim, S., & Choi, C. (2011). Prediction and comparison of urban growth by land suitability index mapping using GIS and RS in South Korea. *Landscape and Urban Planning*, *99*(2), 104-114.

147. Pentland, A. (2020). Diversity of Idea Flows and Economic Growth. *Journal of Social Computing*, *1*(1), 71-81.

148. Pina, K., & Tether, B. S. (2016). Towards understanding variety in knowledge intensive business services by distinguishing their knowledge bases. *Research Policy*, *45*(2), 401-413.

149. Pothukuchi, K. (2005). Attracting supermarkets to inner-city neighborhoods: economic development outside the box. *Economic Development Quarterly*, *19*(3), 232-244.

150. Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, *50*(1), 203-219.

151. Quigley, J. M. (2002). Real estate prices and economic cycles.

152. Rajesh, R., Agariya, A. K., & Rajendran, C. (2021). Predicting resilience in retailing using grey theory and moving probability based Markov models. *Journal of Retailing and Consumer Services*, *62*, 102599.

153. Reades, J., De Souza, J., & Hubbard, P. (2019). Understanding urban gentrification through machine learning. *Urban Studies*, *56*(5), 922-942.

154. Renne, J. L. (2005). Transit-oriented development in Western Australia: attitudes, obstacles and opportunities. *TRANSIT ORIENTED DEVELOPMENT: MAKING IT HAPPEN, 2005, PERTH, WESTERN AUSTRALIA, AUSTRALIA*.

155. Rimal, B., Zhang, L., Keshtkar, H., Haack, B. N., Rijal, S., & Zhang, P. (2018). Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and markov chain. *ISPRS International Journal of Geo-Information*, *7*(4), 154.

156. Ron-Ferguson, N., Chin, J. T., & Kwon, Y. (2021). Leveraging machine learning to understand urban change with net construction. *Landscape and Urban Planning*, *216*, 104239.

157. Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2119-2171). Elsevier.

158. Ryan, S. (1999). Property values and transportation facilities: finding the transportation-land use connection. *Journal of Planning Literature*, *13*(4), 412-427.

159. Salon, D. (2015). Heterogeneity in the relationship between the built environment and driving: Focus on neighborhood type and travel purpose. *Research in Transportation Economics*, *52*, 34-45.

160. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85-117.

161. Schmidt, A., Bardaka, E., & Thill, J. C. (2022). Causal, spatiotemporal impacts of transit investments: Exploring spatial heterogeneity from announcement through long-run operation. *Transportation Research Part A: Policy and Practice, 161*, 151-169.

162. Shafik, N. (1994). Economic development and environmental quality: an econometric analysis. *Oxford Economic Papers*, 757-773.

163. Sinclair, R. (1967). Von Thünen and urban sprawl. *Annals of the Association of American Geographers, 57*(1), 72-87.

164. Smith, B. C. (2009). If you promise to build it, will they come? The interaction between local economic development policy and the real estate market: Evidence from tax increment finance districts. *Real Estate Economics, 37*(2), 209-234.

165. Song, Y., Merlin, L., & Rodriguez, D. (2013). Comparing measures of urban land use mix. *Computers, Environment and Urban Systems, 42*, 1-13.

166. Stevens, D., & Dragićević, S. (2007). A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning and Design*, *34*(4), 708-724.

167. Stevens, M. R. (2017). Does compact development make people drive less?. *Journal of the American Planning Association*, *83*(1), 7-18.

168. Stock, J. H., Watson, M. W., & Wallis, K. F. (2008). *2. A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience* (pp. 95-156). University of Chicago Press.

169. Strauss, J. (2013). Does housing drive state-level job growth? Building permits and consumer expectations forecast a state's economic activity. *Journal of Urban Economics, 73*(1), 77-93.

170. Thill, J.-C. “Research on Urban and Regional Systems: Contributions from GIS&T, Spatial Analysis, and Location Modeling,” in Thill, J.-C. (editor), *Innovations in Urban and Regional Systems Science- Contributions from GIS&T, Spatial Analysis and Location Modeling*, Springer Nature, Cham, Switzerland, 2020, 3-20.

171. Toivonen, M. (2004). Expertise as business: Long-term development and future prospects of knowledge-intensive business services (KIBS).

172. Triantakonstantis, D., & Mountrakis, G. (2012). Urban growth prediction: a review of computational models and human perceptions. *Journal of Geographical Information System, 4(6)*, 555–587.

173. U.S. Census Bureau. (2021a). United States Census Bureau’s County Intercensal Tables. Retrieved from https://www.census.gov/data/tables/time-series/demo/popest/intercensal-2000-2010-counties.html

174. U.S. Census Bureau. (2021b). United States Census Bureau’s quick facts. Retrieved from https://www.census.gov/quickfacts/fact/table/mecklenburgcountynorthcarolina/PST045219

175. U.S. Census Bureau. (2021c). LEHD Origin-Destination Employment Statistics Data (2002-2019). Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program, accessed on 2021 at https://lehd.ces.census.gov/data/#lodes. LODES 7.5

176. U.S. Census Bureau. (2021c). United States Census Bureau’s quick facts. Retrieved from

https://www.census.gov/quickfacts/fact/table/mecklenburgcountynorthcarolina/PST045219

177. United States Census Bureau. (2010). 2010 Census urban and rural classification and urban area criteria. *Geographic Concepts*. Data retrieved from https://www2.census.gov/geo/tiger/TIGER2010/UA/2010/

178. United States Census Bureau. (2017). LEHD Origin-Destination Employment Statistics Data (2002–2015). Data retrieved from https://lehd.ces.census.gov/data/

179. United States Census Bureau. (2021b). Annual Resident Population Estimates for States and Counties: April 1, 2010 to July 1, 2019; April 1, 2020; and July 1, 2020 (CO-EST2020). Retrieved from https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates/2020-evaluation-estimates/2010s-counties-total.html

180. US Census Bureau. (2017). County Business Patterns and Economic Census. Data retrieved from https://www.census.gov/data/tables/2017/econ/susb/2017-susb-annual.html

181. US Department of Energy. (2019). Annual Vehicle Miles Traveled in the United States. Retrieved from https://afdc.energy.gov/data/10315

182. Wang, K. (2013). Causality between built environment and travel behavior: Structural equations model applied to Southern California. *Transportation research record*, *2397*(1), 80-88.

183. Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236-244.

184. Warth, G., Braun, A., Assmann, O., Fleckenstein, K., & Hochschild, V. (2020). Prediction of socio-economic indicators for urban planning using VHR satellite imagery and spatial analysis. *Remote Sensing*, *12*(11), 1730.

185. Wassmer, R. W. (2001). Metropolitan prosperity from major league sports in the CBD: Stadia locations or just strength of the central city? A reply to Arthur C. Nelson. *Economic Development Quarterly*, *15*(3), 266-271.

186. Waxman, A. A. P. (1999). *Utilizing economic theories of retail to revitalize inner-city neighborhood business districts: the case of Uphams Corner Main Street* (Doctoral dissertation, Massachusetts Institute of Technology).

187. Wei, Y. (2018). High-speed rail impact on urban economic growth——Based on grey prediction model. In *2018 3rd International Conference on Humanities Science, Management and Education Technology (HSMET 2018)* (pp. 265-268). Atlantis Press.

188. Weissbourd, R. (2006). Strengthening Communities for Regional Prosperity. *The Living Cities Policy Series*, 1-8.

189. White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and planning A*, *25*(8), 1175-1199.

190. White, R., Engelen, G., & Uljee, I. (1997). The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design*, *24*(3), 323-343.

191. White, R., Engelen, G., & Uljee, I. (2015). *Modeling cities and regions as complex systems: From theory to planning applications*. MIT Press.

192. Wilson, B., & Song, Y. (2009). Comparing apples with apples: how different are recent residential development patterns in Portland and Charlotte?. *Journal of Urbanism*, *2*(1), 51-74.

193. Wirtz, J., Tuzovic, S., & Ehret, M. (2015). Global business services: Increasing specialization and integration of the world economy as drivers of economic growth. *Journal of Service Management*.

194. Wolf-Powers, L. (2005). Up-zoning New York City's mixed-use neighborhoods: Property-led economic development and the anatomy of a planning dilemma. *Journal of Planning Education and Research*, *24*(4), 379-393.

195. World Bank. (2020). Urban development. Retrieved from https://www.worldbank.org/en/topic/urbandevelopment/overview

196. Yang, Q., Li, X., & Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Computers & geosciences*, *34*(6), 592-602.

197. Yang, X., Zheng, X. Q., & Lv, L. N. (2012). A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata. *Ecological Modelling*, *233*, 11-19.

198. Yarbrough, T. S. (2014). *Retail change and light rail: An exploration of business location changes accompanying commuter rail development in Denton County, Texas*. University of North Texas.

199. Zhang, L., Hong, J., Nasri, A., & Shen, Q. (2012). How built environment affects travel behavior: A comparative analysis of the connections between land use and vehicle miles traveled in US cities. *Journal of Transport and Land Use*, *5*(3), 40-52.

200. Zhene, Z., Hao, P., Lin, L., Guixi, X., Du, B., Bhuiyan, M. Z. A., ... & Li, D. (2018, October). Deep convolutional mesh RNN for urban traffic passenger flows prediction. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)* (pp. 1305-1310). IEEE.

201. Zheng, Y., Xu, Z., & Xiao, A. (2023). Deep learning in economics: a systematic and critical review. *Artificial Intelligence Review*, 1-43.

202. Zhou, B., & Kockelman, K. M. (2008). Self-selection in home choice: Use of treatment effects in evaluating relationship between built environment and travel behavior. *Transportation Research Record*, *2077*(1), 54-61.

203. Zoričak, M., Horváth, D., Gazda, V., & Hudec, O. (2021). Spatial evolution of industries modelled by cellular automata. *Journal of Business Research*, *129*, 580-588.

204. Zukin, S., Trujillo, V., Frase, P., Jackson, D., Recuber, T., & Walker, A. (2009). New retail capital and neighborhood change: Boutiques and gentrification in New York City. *City & Community*, *8*(1), 47-64.